Three Essays on Job-Trainee and Employee Behavior: Experimental Evidence from Malawi

by

Susan Godlonton

A dissertation in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Public Policy and Economics)
in the University of Michigan
2013

Doctoral Committee:

Professor Jeffrey Andrew Smith, Co-Chair
Assistant Professor Rebecca Lynn Thornton, Co-Chair
Associate Professor Sarah Andrea Burgard
Professor David A Lam

ACKNOWLEDGEMENTS

There are many people to whom I am greatly indebted for their support, encouragement and mentorship during graduate school. To my advisors who have provided considerable understanding, mentorship and guidance. Rebecca Thornton, you have assisted me every step of the way. You have not only shared your skills and training, but have ensured I have a good understanding of the professional life of an economist and have empowered me every step of the way. Jeff Smith, you have offered exceptional technical guidance. David Lam, you have always provided wisdom and support regardless of the situation. I sincerely appreciate the opportunity to have learnt from you. To my friend and mentor, Jess Goldberg, I am not sure I could have done this without your unfaltering support and mentorship.

I am also grateful to my family for their support. Mom and Jen thank you for coming to Malawi and sharing some fieldwork experiences with me. I would like to extend special thanks to study partners: Yoonsun Han, Yunjung Kim, Caroline Weber and Jess Williams for all their support during the early years. To my research group: Emily Beam, Rob Garlick and Brad Hershbein, for all their valuable input in the later years. To my extraordinary supportive and encouraging officemates: Jiang Jiang, Hyunjee Kim, Yunjung Kim and Ana-Maria Mocanu and Caroline Weber; I am sincerely grateful I spent so much time with you. Hopefully we can carry the enthusiasm and work ethic of Lorch 112 to our new jobs!

There are a number of people who supported me getting to graduate school or more generally helped me along the path to becoming an economist. I am grateful to Justine Burns, Malcolm Keswell, Servaas van der Berg, and Neil Le Roux for the roles they played in my undergraduate and Masters studies. Martin Wittenberg played an extraordinary role as mentor and without his encouragement to start the PhD I would most likely never have taken this path.

My fieldwork in Malawi would not have been possible without the assistance of a very dedicated team. I offer special thanks to Kelvin Balakasi, Ernest Mlenga, Christopher Nyirenda and Denise Matthijisee who have all played critical roles in the field and have overcome many challenges with implementing the projects used in this dissertation.

My research has been supported with grants from the Rackham Graduate School, and the University of Michigan African Studies Center. I gratefully acknowledge use of the services and facilities of the Population Studies Center at the University of Michigan, funded by NICHD Center Grant R24 HD041028.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Three Essays on Job-Trainee and Employee Behavior: Experimental Evidence from Malawi

by

Susan Godlonton

Co-Chairs: Jeffrey Smith and Rebecca Thornton

In the first chapter I examine the relationship between employment risk and job seeker performance. To induce exogenous variation in employment risk, I randomize outside options for job seekers undergoing a real recruitment process. I do this by assigning job seekers a 0, 1, 5, 50, 75 or 100 percent chance of real alternative employment of the same duration and wage as the jobs for which they are applying. I find that job seeker performance is highest and effort is lowest among those assigned the lowest employment risk (a guaranteed alternative job), and performance is lowest and effort highest among those facing the highest employment risk (those without any job guarantee). My findings are consistent with a framework that ties together insights from economics and psychology; performance is an increasing function of effort and an inverse u-shaped function of stress.

In the second chapter I exploit the experiment used in chapter one and estimate the employment and wage effects of a short term job. I find the following key results. First, there is a 10.6 to 13.9 percentage point increase in average employment during the eight months following the job. Second, there is a sizeable increase in wages. Individuals earn approximately 60 to 67 percent more per day. There is suggestive evidence that individuals are switching into

different occupations particularly clerical and related work away from agricultural based activities. Lastly, the estimated returns to the job are larger among those who perform worst on a high stakes numeracy and literacy test.

In the third chapter, examines corrupt behavior by interviewer employees working on short term contracts in a developing country. Specifically, I measure how employees change the extent to which they steal from the firm in response to varying degrees of monitoring in the work force. I find that decreasing the monitoring rate by ten percentage points increases the likelihood of money being stolen by approximately four percentage points; and the amount stolen by between five and eight percent on baseline theft. I also observe the relationship between the monitoring rate and employee crime to be non-linear.

Chapter 1

Employment Risk and Performance

1.1 Introduction

Risk matters. We study behavioral responses either in reaction to the realization of a risky process, or in response to underlying risk, which typically take the form of risk coping or risk mitigation.[1] The same framework characterizes most studies of employment risk: the emphasis is on understanding either the consequences of a job gain or job loss, or on coping strategies for mitigating income uncertainty.[2,3] However, we know little about how risk might affect performance in the job seeking process or ultimate chance of securing employment.[4] Given high and rising employment uncertainty (ILO, 2012) it is important to understand the process through which risk may affect employment, as well as the extent to which risk has heterogeneous effects across job-seekers. In this paper, I

---

[1] An extensive literature documents how individuals account for risk in their decisions related to many domains for example: insurance and contracting (Arrow, 1971; Grossman and Hart, 1983).

[2] The World Development Report (2013) provides an extensive overview of the individual and social consequences of employment. One key strand of literature focuses on the impact of gaining or losing work. Often this empirical work uses an exogenous shock that results in job loss such as plant closures and retrenchments to examine both short term and long term effects on future employment and earnings (Stevens, 1997; Chan and Stevens, 2001; Ruhm (1991; 1994); Topel, 1990; Schoeni and Dardia, 1996; Gregg and Tominey, 2005; Couch, 2001).

[3] A second strand of relevant literature examines risk coping mechanisms and their impacts in the labor market. This literature has examined the roles of unions (Magruder, 2012), unemployment insurance (Gruber, 1997; Green and Riddell, 1993), and informal networks (Burns, et al. 2010; Beaman and Magruder, 2012) and how individuals use these support structures to mitigate risk of unemployment.

[4] As discussed in-depth in Fafchamps (2010), shocks and risk are often used interchangeably despite being distinct. He highlights the lack of research on the impact of any type of risk in the empirical development literature, which has instead focused on the effects of shocks, ignoring the anticipatory nature of the shocks. This is in contrast to older theoretical work that explicitly addresses this and shows that risk aversion should lead to underinvestment and underproduction (Sandmo, 1971).

explicitly examine the relationship between employment risk and job-seeker performance and employment.

Typically the causal effect of employment risk on performance cannot be estimated because of challenges in measurement and identification. First, measuring employment risk is difficult. For example, research examining the relationship between risk and savings use proxies such as variability in household income, variability in expenditures, or, in more recent work, the probability of a job loss (Caroll, 1994; Dynan, 1993; Lusardi, 1998). An appropriate proxy when measuring employment risk might be the probability of a job gain rather than a job loss. Still, none of these proxies provide a direct measure of risk. An alternate approach is to measure decision making in response to experimentally-induced risk in a laboratory setting. However, while such experiments provide useful insights about potential mechanisms, it is often unclear whether evidence from lab-based experiments will translate into real world behavior. This is especially true when trying to predict the behavior of individuals in developing countries, who face far different overall levels of risk and income, and have very different levels of education and experience, than the university students who are typical subjects of laboratory experiments. Second, even if one could directly measure employment risk, it is usually endogenous to the outcome of interest. For example, in the case of job seeker performance, individuals of higher ability are likely to face lower employment risk, yet also perform better on average, making it hard to establish causality. Lastly, while effort and performance are key mechanisms through which risk affects employment, these concepts are also difficult to measure due to self-reporting biases and lack of good quality data.

In this paper, I overcome these challenges by explicitly varying employment risk using a field experiment to examine the impact of employment risk on performance. I randomize job seekers' outside options during a real recruitment process, working in collaboration with a real recruiter offering short term jobs. I randomly assign 268 job seekers a probabilistic chance (0, 1, 5, 50, 75 or 100 percent) of an alternative job. This reduces the downside risk of performing poorly during the recruitment process. For those with a guaranteed outside option, employment risk is zero. To examine the relationship between employment risk and job seeker performance, I utilize both objective and subjective performance assessments from administrative data. To measure effort, I use indicators from both administrative and self-reported data sources.

I find that improving a job seeker's outside option leads to improved performance while effort declines. Job seekers assigned a guaranteed outside option performed approximately 0.45 standard deviations better on recruiter-administered tests of knowledge taught in training than did those who received no outside option. Moreover, I observe that the relationship between risk and performance is highly non-linear. These findings are confirmed using the quality of active participation in job training as a measure of performance. I find higher quality average engagement in training by those assigned high outside options compared to those assigned no outside option. For effort indicators, I find the reverse; that is, I find that job seekers assigned the highest probability of outside option put forth the lowest effort, while those assigned the lowest outside option put forth the most effort. In terms of punctuality, job seekers assigned a guaranteed outside option were 9.3 percentage points more likely ever to arrive late during the three-day training conducted during recruitment, as compared to those assigned no outside option; however the difference is not statistically significant. I do find large, robust, and statistically significant differences in self-reported effort. Individuals assigned a guaranteed outside option spend 25 minutes less per day studying training materials compared to job trainees assigned no outside option. In sum, I find that performance is highest and effort is lowest among those assigned the lowest employment risk, and performance is lowest and effort highest among those facing the highest employment risk. These results are robust to a number of different specification checks including using multiple observations per person, as well as to a host of robustness checks, such as weighted regressions and Lee bounds, which address concerns arising from differential survey non-response.

While this is the first study, to my knowledge, to examine this question in labor markets, my results are consistent with laboratory experiment findings conducted by Ariely et al. (2009).[5] They conducted laboratory experiments among 76 participants in rural India, offering either a high, medium, or low incentive for meeting a performance target on six different games testing concentration, creativity, or motor skills. These performance incentives are in some sense the inverse of the variation in my experiment: while I decrease risk, high-powered incentives increase it. They consistently find that performance in the group assigned the high incentive (400 Indian Rs, equivalent to a month's salary) was lowest. With the exception of one task, differences in performance between the low incentive group (4 Indian Rs per game) and the medium incentive group (40 Rs per game) were not statistically significant.

[5] Psychologists have extensively studied conditions under which increased pressure to perform has resulted in "choking." Seminal work is presented in Baumeister (1984) and Baumeister and Showers (1986). More recently, Beilock (2010) provides a comprehensive review of this literature, covering performance in sports, academic environments, and professional settings.

My findings, then, are consistent with Ariely et al.'s in the sense that performance is negatively correlated with risk. My contributions go beyond affirming this finding, however. I document that the relationship between risk and performance in this context is highly non-linear. I extend the experiment from the risk associated with wage incentives to study employment risk, a distinct though clearly related construct with potentially larger welfare consequences. I also extend the literature from the lab to the field. The variation in risk in laboratory studies is artificial and over windfall income, but in my setting, the variation is over risk in securing real, meaningful employment equivalent to that for which subjects have chosen to apply through a competitive and arduous process. To my knowledge, no evidence in a real-world setting has illustrated the link between risk, performance, and effort, and as Kamenica (2012) notes, it was uncertain whether the previous findings would extend.

Additionally, I collect rich baseline and outcome data in order to incorporate an important strand of the psychology literature that studies the mechanisms through which risk and uncertainty affect behavior. Many previous studies in economics have only identified the reduced-form relationship between uncertainty or risk and performance, though Angelucci et al. (2012) measure cortisol levels in a laboratory study of how stress affects entrepreneurship. The data I collect allow me to rule out alternative mechanisms. There are a number of behavioral theories that are consistent with the key result that individuals facing a lower incentive to perform (having improved outside options) exhibit higher performance. I attempt to shed light on the underlying mechanism for the observed results. I explore stress, gift exchange, nutritional wage, and stereotype threat as potential mechanisms.

The stress mechanism draws on economic and psychological insights. By improving a job seeker's outside options, the incentive to exert effort during recruitment is reduced. Therefore, as outside options increase, effort in the recruitment process should decline, and therefore so too should performance. I refer to this as the *incentive effect*. However, at the same time, improving a job seeker's outside options, also reduces the stress experienced during the job-seeking process, a premise that is supported by psychology and public health literature that finds that uncertain employment prospects are stressful (Feather, 1990; de Witte 1999, 2005; Burgard et al., 2009). This reduction in stress likely has performance implications as Yerkes-Dodson (1908) show that performance has an inverse u-shaped correlation with arousal (stress). Therefore, as stress decreases due to improved outside options, performance could increase or decrease. I refer to this as the *stress effect*. The resulting predictions suggest that, as risk declines, so too should effort, but it is ambiguous whether performance would increase or decrease. The impact of stress on performance is under-studied within economics. The research that does exist focuses on how stress affects performance in professional activities, sports performance, and academic settings.[6,7,8] In fact, in Kamenica's recent (2012) review

[6] In the public health and industrial psychology literatures, stress has been shown to be correlated with performance among nurses, medical doctors, policemen, and teachers (Jamal, 1984; Motowidlo, 1986; Sullivan and Bhagat, 1992; Band and Manuele, 1987).

[7] The literature on sports performance presents relatively mixed results. Primarily, this literature has looked at the probability of scoring penalty kicks in professional soccer. Dohmen (2008) finds that, when the importance of scoring is greatest, individuals tend to score. Apesteguia and Palacios-Huerta (2010) find that players who shoot second in a penalty shoot-out lose the game 60.5 percent of the time. They argue that this is driven by increased pressure to perform, and identification is achieved because the order of the shoot-out is determined randomly from a coin flip. However, Kocher et al. (2011) fail to replicate these findings using an extended dataset. Paserman (2010) examines performance in tennis and sets up a structural model. He finds that individuals would be substantially more likely to win if they could score when it mattered most.

[8] The literature examining high-stakes academic testing also finds mixed evidence. Ors et al. (forthcoming), find that women perform worse than men on a high-stakes entrance exam for an elite university despite higher performance on other low-stakes exams in France. In the education literature more broadly, testing anxiety has been widely observed and studied. Evidence shows that test anxiety can both increase and decrease performance (Sarason and Mandler, 1952; Tryon,1980 provide extensive reviews).

article, he states that “overall, to date there is no compelling empirical evidence that choking plays an important role in any real-world labor market.” My results fill this gap.

There are a number of other behavioral theories that are consistent with the key result that individuals facing a lower incentive to perform (better outside options) exhibit higher performance. Gift exchange is one possibility and would require the performance results to be driven by increased effort. However, I find that individuals assigned a high probability of an outside option exert less effort in studying for the tests during recruitment, suggesting that gift exchange is not the mechanism driving the observed performance results. Second, the nutritional-wage hypothesis might be a possibility. However, I do not observe differences in food expenditures by treatment group during the training, so it is unlikely that this is the driving mechanism. Third, stereotype threat might be the driving mechanism. However, I find that job trainees’ perceptions about their own likelihood of being hired by the recruiter do not significantly differ across treatment groups, suggesting that stereotype threat is an unlikely mechanism. While my results are consistent with a stress response, I cannot rule out that there is some other psychological consideration that operates in a similar way to stress. Moreover, I cannot identify the mechanism through which stress might act to impair performance.

In my study, the finding that performance is highest among individuals with guaranteed outside options has important policy implications. In this study, differences in employment rates by treatment status show that individuals assigned a 75 or 100 percent chance of alternative jobs were twice as likely to be employed by the recruiter compared to those in the other treatment groups. I also examine heterogeneity of the employment impacts by mental health status and ability. I find no differential employment effects by mental health status. I do find suggestive evidence that the reduction in employment risk has the greatest impact on individuals in the middle of the ability distribution.

Perhaps the broader implications of these results are that individuals with greater income support through employment guarantees, cash transfer programs, family support, or employment income are likely to perform better. This may have positive feedback effects. Poor initial employment probabilities can induce stress-related performance reductions resulting in poverty persistence across individuals, communities, or countries. Lastly, the results yield insights into what types of people are more likely to be hired with different recruitment strategies. For example, individuals exposed to higher employment risk have a greater chance of employment in hiring processes that place greater emphasis on effort than on performance.

There are some limitations to my findings. First, this study was conducted using short term job opportunities; the effects of longer term job security cannot be assessed in this context. Second, the experiment was conducted among a sample of relatively well-educated men in the capital city of Malawi. This paper cannot speak to how other groups might respond. Third, it would have been better to have biomarker indicators to measure stress (e.g. cortisol) directly but due to logistical and budgetary limitations this was not feasible. Fourth, while I do examine the heterogeneity of the performance and employment results and find that risk matters most for those in the middle of the ability distribution and has limited differential effects by mental health status, my ability to draw robust conclusions about heterogeneous effects is limited by my sample size.

The remainder of the paper proceeds as follows: Section 1.2 provides contextual information about labor markets and recruitment in Malawi and presents the experimental design. Section 1.3 outlines the different data sources used. Section 1.4 presents the estimation strategy, and Section 1.5 presents and discusses the results. Section 1.6 concludes.

1.2 Experimental Design

To examine the relationship between employment risk and job trainee performance, I vary individuals' outside options during a real recruitment process. In the absence of this intervention, the distribution of job seekers' outside options is correlated with their own ability, prior work experience, and social networks. I offer job trainees a randomly assigned probability of an alternative job with the same wage and duration as the job for which they have applied. I work in collaboration with a real recruiter and embed the experimental component into an already existing recruitment process. In this section, I provide some background about the experimental setting, outline the details and timeline of the recruitment process, and provide details of the intervention.

1.2.1 Setting

Developing country urban labor markets are characterized by high unemployment and underemployment, as well as high job instability (WDR, 2013). In many respects, these labor markets are similar to low-income labor markets in developed economies. High rates of in-migration to urban areas in developing countries suggest these problems are likely to increase and that rural labor markets are worse. Malawi, the fourth-fastest urbanizing country in Africa (HDR, 2009), is no exception. Data from the nationally representative Integrated Household Survey shows that only 39.8 percent of urban Malawian men aged 18-49 report ever being employed for a wage, salary, or commission in the last 12 months. When examining activity in the last seven days, 29.6 percent report engaging in household agricultural activities; running or helping to run household small businesses; engaging in day labor (known as "ganyu"); or being employed for a wage, salary, or commission. Incidence of job turnover and the prevalence of short term contracting are not well measured. However, sectors that are characterized by high turnover, fixed term contracts, and seasonality are the most common among urban residents. For example, 7.9 percent report working in construction and 46.8 percent in community, social, and personnel services (IHS2010/11).[9]

Due to the recruiter's eligibility restrictions, the sample in this paper is restricted to men aged 18 and older who had completed secondary schooling. Approximately 39 percent of urban men aged 18 to 49 have completed secondary schooling in Malawi. However, they too face high rates of unemployment: only 52.5 percent had worked in the past year (IHS, 2010/11).[10] Due to their relative higher social status, these men also bear considerable financial responsibility not only for their immediate families but often for extended family members. On average, these men report sending 10 percent of their wage income to other households (IHS2010/11).

1.2.2 Recruitment Process and Timeline

The sample of respondents is drawn from a recruitment process hiring interviewers for a health survey.[11] Contract work on survey projects for government or international organizations, research projects, or NGOs is quite common in the capital city. Data collected by Chinkhumba et al. (2012) which samples approximately 1200 men

---

[9] The community, social, and personnel services sector also includes teachers, whom I have excluded when calculating the fraction working in this sector because teaching, while low-paying, is a stable profession in this context.

[10] When examining responses regarding activities in the past 7 days among men with completed secondary school resident in urban areas in the IHS2010/11 data, 1 percent report working in household agricultural activities; 6.2 percent had run or were helping to run small household businesses; 1.95 percent were engaged in ganyu/day labor; and 21.7 percent had been employed for a wage, salary, or commission.

[11] The recruiter conducts independent consulting within Malawi and has, for several years, implemented various randomized controlled trials and other data collection efforts in Malawi for universities and other international NGOs.

aged 18 to 40 in the Malawian capital finds that one in ten individuals had ever worked as an interviewer; of those who had completed secondary schooling, the number was one in four.[12] This data set also provides some descriptive data on hiring practices.[13] A total of 23 percent report having taken a test for their most recently held job. Approximately half (51.5 percent) report being interviewed for their most recently held job. One third report attending job-specific training for their most recently held job.[14]

The jobs offered by the recruiter are relatively high paying, offering approximately three times the average wage for men who have completed secondary school.[15] However, the wages offered by this recruiter are comparable to those offered by other employers hiring for this type of work.[16]

The recruitment process timeline is presented in Figure 1.1. There are three phases of the recruitment process: pre-screening; training and screening; and final selection. The experimental component was conducted during the training and screening phase.

---

[12] These numbers are high and deserve explanation. First, the census of Malawi took place in 2010. Many individuals are likely to have worked for the census as the National Statistics Office hired extensively. Second, 65 percent of individuals only report working once as an interviewer. Third, interviewer is likely broadly interpreted including work individuals may have conducted as market research or other non-research that involved interviewing others.

[13] Unfortunately the Integrated Household Survey, which would provide nationally representative data, asks only a single question related to job search. Individuals who had not worked in the past 7 days are asked whether or not they looked for work in the past four weeks. Moreover, firm level data on hiring practices is not available.

[14] These numbers come from the authors' own tabulations from unpublished data collected by Chinkhumba, Godlonton, and Thornton (2012).

[15] The mandated monthly minimum wage at this time for urban individuals was only \$24 per month. However, more relevant wage information regarding comparable wages can be obtained using the Integrated Household Survey (2010/11). The average wage among 18 to 49 year old urban men who had completed secondary schooling is approximately \$4.75 per day, the median is somewhat lower at \$2.02.

[16] Wages at institutions hiring interviewers regularly (such as Innovations for Poverty Action, the National Statistics Office and others) ranged from \$15 to \$32 per day for urban interviewers. Wages offered in this case are on the low end for this type of work at \$15.

Recruitment Process: Pre-screening

To advertise positions, advertisements are placed in multiple public places.[17] The placement was determined and conducted by the recruiter and followed their standard protocol. The public advertisements for the job included eligibility requirements and the application procedure. To apply, each individual was required to take a pre-screening assessment test and submit a copy of his resume.[18] The written assessment included numeracy and literacy modules and a brief background module. A total of 554 applicants took this written assessment test. Based on the numeracy and literacy scores, the recruiter selected the top 278 applicants to advance to the job training phase of the recruitment process. Individuals selected were screened based on a clear cut-off using the numeracy and literacy test administered. The distributions of these scores are presented in Figures 1.2.a, 1.2.b, and 1.2.c. Given this selection criterion, the sample of interest is a non-representative sample of applicants. However, it is representative of the individuals who were selected for training by the recruiter and therefore captures the population of interest relevant to the research question.

Recruitment Process: Training

The 278 job seekers who advanced to job training attended a pre-training information session. During this session, job trainees were provided with materials required for training and logistical information related to the training process. They were also informed about the opportunity to participate in this research study. A total of 268 applicants of the applicant pool opted to participate (95 percent). This constitutes the main sample. Consenting participants were asked to self-administer a baseline

[17] These include public libraries, educational institutions, public notice boards, and along streets.
[18] Individuals were encouraged to bring their resumes. Most (95 percent) did bring a resume. Those who did not bring a resume were not prevented from taking the pre-screening assessment test.

questionnaire after which they were issued their randomly assigned probabilistic job guarantee. Details related to the nature and assignment of the probabilistic job guarantees are discussed in Section 1.2.3.

All 278 job trainees were invited to attend three days of full-time training and further screening. They were paid a wage equivalent to half of the daily wage of the employment opportunity for each day of training attended. During training, applicants were monitored for their punctuality and the quantity and quality of active participation in the job training in which they learned about the health survey for which they were being trained. Individuals were also tested on materials taught. Summary statistics and details related to these administrative data are discussed in Section 1.3.

Also, for the purposes of this study, on each day of training, respondents were asked to self-administer a survey questionnaire. The recruitment team did not know who chose to participate in the research, what alternative job probabilities were assigned, or which participants completed the daily questionnaires. Moreover, the recruiter did not get access to the survey questionnaires. This was carefully explained to the respondents and monitored to ensure confidentiality regarding participation in the research study.

At the end of the final day of training, the alternative job draws were conducted, and participants learned their alternative job employment realization. The recruiting team was not present at this time, and they were not at any point informed as to who received an outside job offer.

Recruitment Process: Selection by Recruiter

Two days after completing the training, the successful applicants for the job advertised by the recruiter were contacted.

1.2.3 Intervention: Probabilistic Outside Employment Options

During the information session prior to the commencement of the job training, job trainees were randomly assigned some *probability of employment* via a job guarantee for an alternative job. There were six different probabilistic guarantees – 0, 1, 5, 50, 75, and 100 percent chance of an alternative job.[19] Thus, the intervention experimentally altered individuals' outside options.

The alternative jobs were constructed to mimic as closely as possible the jobs offered by the recruiter. The alternative jobs were for the same duration and pay as the job being offered by the recruiter. They were real jobs, requiring real effort and paying real wages. While the recruiter was hiring for interviewer positions, the alternative jobs were other research jobs. In both cases, individuals were working for research projects for the same university albeit on different projects and performing different types of research tasks. The alternative jobs included data entry, translation, transcription, and archival research.[20]

Individual treatment status was blind to the research and recruitment team but known to the job trainee. Each job trainee was given an envelope with his employment ID written on it. Inside the sealed envelope was an employment contract stating which probabilistic job guarantee he had received.[21] Job trainees assigned a 0 percent chance of an alternative job also received an envelope. Randomization was conducted at an individual level and stratified on quintiles of baseline ability and an indicator variable for

[19] In a pilot version of this experiment, there also existed a 25 percent chance of a job guarantee. However, given the results of the pilot, the sample size required to detect reasonable effect sizes was larger than the financial constraints of this project would allow. While I would have liked to have included a 99 percent chance of a job guarantee to test differences in small changes in risk at different points in the distribution (specifically 0 to 1 percent and 99 to 100 percent) due to budgetary limitations, it was not possible to implement this. I hope to explore this in future work.
[20] If individuals were selected by the recruiter, and also received an alternative job they were required to take the recruiter's job and not the alternative job.
[21] Individuals could choose to reveal their contract to anyone within or outside of the group but they were not required nor encouraged to do so.

whether or not they had ever worked for the recruiter.[22] Baseline ability was determined using participants' scores from the numeracy and literacy components of the pre-training assessment test. The distribution of the probabilities was pre-assigned to the 278 applicants invited to attend the training stage of the recruitment process. Ten individuals opted not to participate in the research project or in the recruitment process. These participants made their decision before knowing to which treatment group they had been assigned. In the final sample of the 268 male participants, the distribution of the probabilistic job guarantees is similar to the intended assignment (Table 1.1, Panel A).

The distribution of treatment allocated approximately 20 percent of the sample to each of the 0, 1, 5, and 50 percent chance groups and approximately 10 percent of the sample to each of the 75 and 100 percent chance groups.[23] Respondents were informed about the distribution of the alternative job probabilities prior to learning their own treatment assignment, so as to ensure that all participants had the same beliefs about the distribution. Had respondents not been told the underlying distribution, then individuals would have variable information, which would be endogenous to the truthfulness, candor, and size of their social network among other job trainees.

Job trainees were also informed that their treatment assignment would not be revealed by the research team to the recruitment team or anyone else. It was consistently emphasized that their probability of an alternative job would have no direct bearing on their probability of being hired by the recruiter. To ensure individuals were clearly informed about how the probabilities worked and how the draws would be conducted, they were discussed in detail and demonstrations were conducted to illustrate the process.

[22] "Ever worked for the recruiter" is broadly defined. That is, even individuals who had attended a prior job training session held by the recruiter but had never successfully been employed are included in this category.
[23] While equal proportions across groups was desirable, this was not feasible due to budgetary limitations.

The draws were conducted in the following way: if a job applicant received a probabilistic job guarantee of 75 percent, then after the conclusion of the final day of training, he faced a bag of 100 bottle tops. In the bag, there were 75 red bottle tops and 25 green bottle tops. If the individual drew a red bottle top, he would receive an alternative job; if he selected a green bottle top, he would not. Corresponding procedures were used for the other treatment groups.

An important concern is whether individuals actually understood the probabilistic nature of the alternative job offers.[24] After the treatment was explained, but before individuals learned their own probability, we surveyed participants to elicit their perceptions related to their understanding of these probabilities. Participants were asked for each treatment arm what they expected the realization of alternative jobs to be. For example: "If 60 participants received the 50 percent job guarantee, how many of them are likely to receive an *alternative* job." The modal response by participants was fairly accurate. For the 5 and 50 percent treatment groups the modal response translated into 5 and 50 percent respectively. For the 1 percent group the modal response was 1.6 percent.[25] For the 75 percent treatment group the modal response translated into an 83 percent chance, which is a slight overestimate. In general, it seems reasonable to assume that participants understood the assigned outside options.[26, 27]

[24] Although the sample is relatively well-educated, mathematical literacy, particularly related to probability, is not universal. For example, one of the numeracy questions during the selection screening test asks: "To pass an exam which comprises a part A, B, and C, a person needs to pass not less than 40 percent in A, not less than 30 percent in B and not less than 30 percent in C. If A, B, and C have 50, 30 and 20 marks respectively, what is the minimum mark to pass the exam?" Only 45 percent of the sample of job trainees answered this question correctly.

[25] Given the phrasing of the question for the 1 percent chance treatment group, it was impossible for individuals to select an integer that would map into 1 percent of the distribution getting alternative jobs. The modal response was 1 person, which maps to the 1.6 percent. The second most frequent response recorded was 0.

[26] These prior perceptions are not differential across treatment status.

[27] Open ended questions on the survey asked respondents to explain how they understood the job probabilities. The responses in general suggest they understood how this worked. For example: "The probability criteria are dependent on the chance and not merit of a person in terms of experience and qualification."; "Those that have 75% chance have higher chances as compared to those that have 1% chance."; "It's a good idea after all if you are guaranteed a 100%

1.3 Data

I use two sources of data in this paper. Primarily, I use administrative data collected by the recruiter. I supplement this with survey data I collected for this project using respondents' self-administered questionnaires.

1.3.1 Baseline Data

Pre-screening assessment test (administrative data): From the recruiter, I have data from the pre-screening assessment test conducted to select the job trainees. This test consists of numeracy, literacy, and background information modules.[28] Among the 554 job applicants, the average numeracy score was 52.5 percent, and the average literacy score was 70.3 percent. For the sample of short-listed candidates, the sample frame for this paper, the average numeracy score was 63 percent, and the literacy score was 80 percent. The ability score that will be referred to throughout the remainder of the paper is a composite measure of the individual's numeracy and literacy scores equal to the sum of the two scores. The distribution of the numeracy, literacy, and composite ability scores are presented in Figures 2a, 2b and 2c.

probability you don't have to worry about the other job." "Those who have 100%, 75%, 50% have a high chance of getting an alternative job whilst those who have a 5% and 1% have a low chance."

[28] During the pilot, a similar standardized test of literacy and numeracy was used, but the literacy component was slightly too easy, and this was adjusted for the population in this implementation. The literacy module comprises questions taken from the South African Cape Area Panel Study Wave 1 survey and is supplemented with additional more difficult literacy questions. A large proportion of the numeracy module used comes from the South African National Income and Dynamics Survey wave 1 survey. Additional questions come from previous recruitment tests used by the recruiter as well as other survey implementers in the country, such as the Malawi National Statistics Office.

Baseline questionnaire (survey data): To supplement this administrative data, I conducted a baseline questionnaire. The survey instrument was self-administered during the information session to consenting participants before the commencement of training. It includes questions about previous work experience, employment perceptions and attitudes, physical and mental health indicators, time use, and a work and health retrospective calendar history.

1.3.2 Training and Post-Training Data

I use administrative data collected during the training as well as the hiring decisions made by the recruiter to construct the key outcome variables of interest used in the analysis. I supplement this with daily follow-up survey questionnaires that were also conducted after each day's training.

Participation in training: Table 1.1 Panel B presents the participation rates of the 268 consenting participants. Most of the selected job trainees opted to participate in the training – 94 percent attended training every day. There is no statistically significant difference in training participation across treatment groups.

Punctuality records: Recruitment staff recorded daily attendance including job trainee arrival times. Participants were required to sign in when they arrived to determine to which classroom they had been assigned for the day. When participants signed in, their arrival times were recorded. I use the sign-in times to measure punctuality as an effort indicator.

Room assignment: Participants were randomly assigned to one of three training rooms on day one. On day two they were randomly assigned one of the other two rooms, and on the third day they were assigned to the remaining room.[29]

Test scores: On each training day, a test was administered to job trainees by the recruiter. These test trainees' comprehension of the materials taught during the training sessions and are the most important observable performance indicator used by the recruiter in making employment decisions. Refer to Appendix A for a detailed discussion on the determinants of hiring decisions.

Contribution records: Recruitment staff also recorded the verbal contributions made by job trainees. These records enable me to construct a performance indicator of engagement. Similar measures of engagement have been used in the education literature, typically in the context of teacher evaluations of student engagement (e.g. Dee and West, 2011; Friedricks et al. 2004 reviews the education literature pertaining to student engagement). I also construct a subjective assessment of the quality of the contribution made. The quality scale is graded as *Good, Neutral* or *Bad*. In some cases, multiple members of the recruitment team were documenting these contributions. To eliminate double counting, I count a contribution only once assuming that it came within five minutes of a second contribution. In instances where a contribution is recorded twice and the two records differ, I use the lowest quality assessment. The double counting allows

[29] This ensures that all participants were in a different room on each training day. Although the same materials were taught simultaneously across training rooms, the recruiter felt it was necessary for the participants to be exposed to all the different trainers. All three training rooms were at the same venue. Participants were free to sit as they desired within the assigned room; their seating choice was recorded by the recruitment team. These records are used in later analysis.

me to assess the correlation in subjective assessments made. In 61.5 percent of cases, the two separate records were in agreement.[30]

Employment records: I obtain the employment records of the consenting job trainees. For each trainee, I have a binary indicator of whether or not he was offered a contract by the recruiter.

Daily survey questionnaires: I supplement these administrative data sources with daily self-administered follow-up questionnaires. While respondents were completing these surveys in private, all recruitment staff left the training venue. Research staff were available to address any questions. Participants were asked to drop their completed questionnaires in a sealed drop box at the venue. The daily questionnaire asked about time use and mental and physical health, as well as employment attitudes and beliefs.

Table 1.1 Panel B presents survey data completion rates. There is some evidence of differential non-response with the follow-up questionnaires by treatment status.[31] Only 83 percent of the participants who received no chance of an alternative job completed the follow-up survey questionnaire every day compared to 96 percent among those who were assigned a 100 percent chance of an alternative job. This difference of 13 percentage points is significant at the five percent level.[32] I primarily use the follow-up data to examine the impact of the outside options on self-reported effort as well as to shed light on the potential mechanism driving the performance results. To address the differential non-response in the survey data, I conduct a number of specification checks that are

[30] Additionally, in 26.5 percent of cases, one record reports the contribution as *good* while the other rates it as *neutral*. In 9.64 percent of cases, one record reports the contribution as *neutral* and the other rates it as *bad*. Finally, in only two cases in which the quality assessments differ does one report assess it as *good* and the other as *bad*.

[31] Completing the daily questionnaires was not a condition of receiving the alternative job.

[32] Differential completion rates are largest on day 1 and decline across time. By day 3, there is no differential attrition across treatment status for the follow-up survey completion. One possibility is that any resentment towards the research project due to being assigned a low probability of an alternative job declined over time. This is consistent with the happiness literature that shows that shocks to happiness are mitigated across time (Kimball, 2006).

discussed in Section 1.4. The results are presented in Section 1.5.3 and in general show that the results are robust to these checks.

1.3.3 Sample

The sample used in the analysis in this paper comprises the 268 consenting job trainees. Table 1.2 presents summary statistics about the sample. On average, respondents are 25 years old, and 18 percent are married. Approximately 17.6 percent of the men in the sample had at least one child. Because of the recruiter's secondary schooling requirement, respondents are relatively well educated for Malawi, with an average of 13 years of education.[33] Respondents report earnings of approximately $220 over the last three months.

Most of the men, 86.9 percent, report having worked previously. Although most men (86.1 percent) had worked at some point during the previous six months, they had only worked, on average, 2.7 months of the preceding six months.[34] Individuals who had previously worked were asked a series of questions about their three most recent jobs. For their most recent job, 58 percent report competing for it, 26.8 percent were required to take a test as part of the hiring process, almost 70 percent were required to attend an interview, and slightly more than half were required to complete some job training prior to employment.[35] In sum, the process is not atypical of hiring processes in this context.

1.4 Estimation Strategy

[33] Although this is relatively high for Malawi in general, it is not atypical for a representative sample of men in urban Malawi. In another survey (Chinkhumba, et al. 2012) that randomly selected men, the average was 11 years of schooling.

[34] The sample used here is similar to the nationally representative integrated household survey sample in terms of key work-related characteristics. For instance, respondents in the IHS10/11 had worked, on average, 5.6 months out of the previous 12 months.

[35] Averages across the three most recent jobs are similar (results not shown).

In this section, I discuss the key outcome variables of interest followed by the main estimating equation. I then discuss the validity of the random assignment in the sample. Lastly, I briefly discuss key alternative specifications implemented as robustness checks.

1.4.1 Key outcomes

To measure performance, I rely on administrative records only. I use test scores from the training assessment tests, as well as engagement in training. I use both quantity and quality measures of engagement in training: cumulative number of contributions; total number of good contributions; total number of neutral contributions; and total number of bad contributions. I construct a performance index measure as a summary index of these performance indicators. The index is constructed as the average of the normalized values of each of these measures (Kling, Liebman and Katz, 2007).

To measure effort, I use both administrative data and survey data. From the administrative data, I use the rich arrival data and construct measures of punctuality: ever late; always late; and how early or late. Using the survey data, I use the self-reported time use diaries to measure the average number of hours per day spent studying training materials and the number spent on leisure activities (watching television, listening to radio). I construct an effort index as a summary measure of effort using the average of the normalized values of the minutes arrived late and time use variables.

Lastly, I examine employment outcomes using data from the recruiter on which respondents were hired.

1.4.2 Main empirical specification

The experimental design of the study permits a relatively straightforward analysis. To estimate the differential performance, effort, and employment by treatment group, I estimate the following regression:

$$Y_i = \beta_1 T0_i + \beta_2 T1_i + \beta_3 T5_i + \beta_4 T50_i + \beta_5 T75_i + \beta_6 T100_i + X_i{}'\beta + \varepsilon_i \qquad (1)$$

where: Y_i indicates job trainee *i's average* performance or effort. The average for each indicator is constructed using data from three observations per individual. In the case of missing data, the average is constructed from the observations available. The indicators *T0*, *T1*, *T5*, *T50*, *T75* and *T100* are binary variables equal to one if the individual received a 0, 1, 5, 50, 75 or 100 percent chance of an alternative job, respectively, and zero otherwise. Rather than assuming a linear relationship, I specifically allow a flexible non-linear relationship between the probabilistic job guarantees and the outcome variables of interest. This allows me to examine the reduced form relationship between employment risk and performance and effort.

Lastly, X_i is a vector of covariates including stratification cell fixed effects, ability score, previous experience with the recruiter, age, and other background characteristics. To facilitate easier interpretation of the coefficients, I demean all control variables, so coefficients are interpretable as group means at the mean of all controls in the regression. Unlike many program evaluation randomized controlled trials, there is no clear control group in my sample. Although being offered no outside option is akin to what individuals would face in the absence of this experiment, this is not a clean control group as these individuals are allocated a poor draw for the purposes of the research.

My main comparison of interest is between those assigned no outside option (*T0*), and those assigned the alternative employment guarantee (*T100*), removing all risk from the job application process. While employment risk is decreasing in the magnitude of the outside option, uncertainty of the alternative jobs is highest among those in the 50 percent group. I do, however, present the average performance, effort, and employment results for all treatment groups, yielding insights into the relationship of these outcomes across the distribution of the outside options assigned.

Given the random assignment of individuals to the different treatment groups, the identification assumption that assignment to treatment group is orthogonal to the error term should hold. One test of this assumption is to compare observable characteristics across the different treatment arms. Table 1.2 shows that the different treatment arms appear to be balanced when examining multiple baseline characteristics. In most cases, I cannot reject the null hypothesis that the means are jointly equal across all the treatments. Similarly, for most pair-wise comparisons, I cannot reject that equality of the means. As assignment was predetermined, no strategic behavior to change treatment status was possible. Controls will be included in the results that follow, but the results are robust to whether or not controls are included.

1.4.3 Alternative specifications

I conduct a host of robustness checks. First, for binary performance or effort indicators, I use probit specifications. Second, to address any remaining concerns regarding imbalance of treatment assignment, I present the analysis with and without controls; I also construct a measure of the extent to which omitted variable bias would have to differ in unobservables relative to observables to explain away the observed differences in performance and effort by treatment group (Altonji et al., 2005; Bellows and Miguel, 2008). Third, I address missing data in the administrative records and differential non-response in survey data records. I use three strategies to address both of these concerns. I follow Fitzgerald, Gottschalk and Moffitt (1998) and present weighted results. I then present conservative bounded results where I implement min-max bounds (Horowitz and Manski, 1998). Lastly, I restrict the sample to the 0 and 100 percent treatment groups and estimate Lee (2009) bounds on the average treatment effect of the 100 percent group relative to the 0 percent treatment group. I discuss the implications of each of these robustness checks for the performance and effort indicators in Section 1.5.3.

1.5 Results

First, I present the performance results using administrative data including training test scores and measures of engagement in training. Second, I present and discuss the effort results using indicators from both administrative data (e.g. punctuality) and self-reported data (e.g. time spent studying training materials). Third, I present a broad set of robustness checks for the performance and effort results. Fourth, I discuss potential mechanisms that may be driving the performance and effort results. Fifth, I present the welfare implications of employment risk by examining differences in employment by treatment group. Lastly, I discuss heterogeneity of the performance and employment results by baseline mental health status and ability.

1.5.1 Performance Indicators

I use two key indicators of performance in the analysis: job trainees' performance on administered tests and engagement in the training. To measure engagement in training, I examine differences across treatment groups in the quantity and quality of verbal contributions.

Administrative Training Tests

The most important assessment tool used by the recruiter for hiring decisions is the performance of the job trainees on the written tests administered during the job training. The correlation between performance on these tests and the probability of being hired by the recruiter is 0.60. The R-squared of a univariate regression of employment on the standardized average test score is 0.357, and the coefficient in this case is 0.225 (standard error is 0.0311). Therefore, for every additional standard deviation in test score, an individual is 31 percentage points more likely to be hired. The determinants of hiring are presented in Appendix Table A.1 and discussed in detail in Appendix A.

Figure 1.3 and Table 1.3 present the main test results using the average performance on the standardized test scores as the dependent variable. I find that job trainees assigned no outside option performed significantly worse than those assigned a 100 percent outside option. The magnitude of the difference ranges from 0.438 to 0.451 standard deviations depending on the set of controls used and is consistently significant at the 10 percent level. The magnitude of these effect sizes is quite large. Perhaps the best way of contextualizing the effects is to compare them to education interventions that aim to impact test scores in developing countries. Kremer and Holla (2008) review randomized controlled trials of education interventions conducted in developing countries. Test score effect sizes from the 26 papers reviewed range between zero and 0.46 standard deviations, with the exception of a technology assisted education intervention in Nicaragua that found large effects of 1.5 standard deviations (Heyneman, 1981). The median effect size from this review was 0.16 standard deviations.

Figure 1.3 and Table 1.3 also show suggestive evidence in support of an increasing trend of performance as a function of employment risk. One exception to this trend is the relatively poor performance of those assigned the 75 percent chance of an alternative job. There is considerable variation in the performance of this group, which I explore further in Section 1.5.6 when examining heterogeneity of the impacts. Moreover, the results show that there are substantial non-linearities in the performance-risk relationship.

Verbal Contributions

Next, I examine differential performance across treatment groups for verbal contributions made during the training sessions monitored by the recruitment team. Appendix A highlights the importance of good quality engagement during training as it is a key predictor of employment in the current context.[36]

I construct both a quantity and quality measure of trainee engagement. More than half of the participants (67 percent) made a contribution at least once during the course of training. Individuals who contributed did so an average of 2.3 times. Approximately 46 percent of the contributions made were classified as *good*, 39 percent as *neutral* and 15 percent as *bad*.

Table 1.4 presents the regression results that control for covariates and stratification cell fixed effects, although the results are robust to excluding covariates (Appendix Table B.3). The performance indicators used here aggregate performance

[36] Classroom behavior in schools has also been shown to be important for labor market success (Segal, 2008, 2012, and forthcoming). I do have a similar measure of behavior to that used in this literature. However, in my setting, training classroom behavior was not an important predictor for determining employment outcomes (See Appendix A). Recruitment staff recorded disruptions by participants during the training sessions. Disruptions include answering phone calls, exiting and re-entering the room, making jokes, and chatting to other trainees, among other things. Almost half of the participants (47.1 percent) were recorded as being disruptive at some point during the training. The total average number of disruptions made was 2.11, conditional on making any disruption. In 47 percent of the cases, the disruptive behavior relates to making noise, chatting with friends, banging on desks etc; in 42 percent of the cases, the disruptive behavior relates to unnecessary moving around the room, or entering and exiting the training room; and in 11 percent of cases relates to participants answering cell phones during training. Using this data, I construct measures of whether the job trainee was ever disruptive, the number of times he was disruptive and the number of each type of disruption. I do not observe statistically significant differences across treatment groups (See Appendix Table B.2).

across the three training days. Column 1 shows that job trainees receiving the 100 percent chance of the alternative job were 11.2 percentage points more likely to make any type of verbal contribution. Probit results are broadly consistent. While these differences are quantitatively large, they are not statistically significant. The total number of contributions is an alternative measure of the quantity of engagement. Table 1.4 Column 2 shows that job seekers assigned a guaranteed outside option make 0.744 more contributions than those assigned no outside option.

In determining employment decisions, a key dimension is the subjective quality assessment of engagement. Appendix A shows that making a *good* contribution impacts the probability of being hired. Participants receiving the 100 percent job probability make 0.410 additional *good* contributions relative to those in the 0 percent job probability treatment group. This difference is statistically significant at the 10 percent level. In fact, individuals in the 0 percent group are the least likely of all groups to make a *good* contribution (only 0.528 contributions on average).[37] This is consistent with the test performance results, which showed that individuals in the 0 percent group performed the worst on average, and those in the 100 percent treatment group performed the best (Table 1.4, Column 3). Similarly, job trainees assigned to the 100 percent treatment group are the most likely to make *neutral* contributions, but the difference is not statistically significant (p=0.127).

Performance index

To address the issue of multiple inferences, I create a performance index. This index is the mean of the normalized value of the average test score and all the verbal

[37] This also goes against a story where the zero probability group is more risk averse and as a result set a higher bar on their ex-ante beliefs about contribution quality. This implies fewer but higher average quality of contributions from the zero probability group.

engagement measures (Kling, Liebman and Katz, 2007). Table 1.4 Column 6 presents these results. Individuals assigned no outside option perform 0.369 standard deviations worse than those assigned the guaranteed outside option. The difference is statistically significant at the five percent level. It is also interesting to look beyond the mean and consider the performance index distribution. Figure 1.4 presents the distribution of this index for individuals assigned no outside option (T0) and those assigned the employment guarantee (T100). This figure shows that the performance distribution for those guaranteed outside employment is shifted quite significantly to the right. The p-value associated with a Kolmogorov-Smirnov distribution test of equality is 0.043. This figure shows that the average performance result is not driven by outliers but rather a shift in the distribution.

In sum, I find that performance is highest among those assigned a guaranteed outside option and lowest among those assigned the lowest outside options. Differences are large in magnitude and often statistically significant. This suggests that reducing employment risk can, at least in this context, result in overall higher performance. Next, I examine effort indicators to assess whether these results are driven by changes in effort.

1.5.2 Effort Indicators

To measure effort, I use administrative data to measure punctuality, as well as self-reported data to observe time use. I also combine these data to construct an effort index as a summary measure of effort.

Punctuality

One potentially important indicator of effort is punctuality. On average, job trainees arrived 21 minutes prior to the beginning of the training start time. Approximately 16 percent arrived late on the first day, 11 percent on the second, and only five percent on the final day (results not shown). Evidently, job trainees realized that their punctuality was being recorded and altered their behavior over time.[38]

To measure punctuality, I use three measures: ever late, always late, and average minutes early/late across the three training days. Table 1.5 shows that individuals assigned to the 100 percent treatment group are 9.3 percentage points more likely to ever arrive late and 6.3 percentage points more likely to be always late compared to those assigned no outside option. These difference are large in magnitude but are not statistically significant (p=0.34; p=0.271). Probit results are broadly consistent.

Table 1.5 Column 3 presents average minutes arrived early or late. I do not observe statistically significant differences in arrival times. To explore this further, I use Kolmogorov-Smirnov distribution tests of equality. I cannot reject at any reasonable level of significance that the distributions of arrival times on each day comparing any two treatment groups are the same (Appendix Table B.2).

Time use

A second dimension is self-reported effort. As part of the daily follow-up questionnaires, individuals report activities in a time use module. I focus on two key categories: time spent studying training materials and leisure time spent listening to the radio or watching television. I construct a measure of the average number of hours spent on each of these activities per training day.

Table 1.5 Column 4 presents the mean number of hours spent studying the training materials for each treatment group. Those with the guarantee of employment report

[38] An alternative explanation is that individuals learned across time how long it would take them to get to the venue as most relied on public transportation that can be very unreliable.

spending the least amount of time studying the training materials, as much as 25 minutes less per day than those who received no chance of alternative employment.[39] Moreover, Table 1.5 Column 5 indicates that individuals in the 100 percent chance treatment group spent 53 more minutes watching television or listening to the radio than those assigned no outside option. These results suggest individuals are substituting time spent studying the training materials for leisure time.

Effort index

I create an effort index similar to the performance index. As with the performance index, this index is the mean of the normalized value of the average minutes arrived early or late; number of hours spent studying the training materials; and number of hours spent watching television and listening to the radio (Kling, Liebman and Katz, 2007). The results are presented in Table 1.5 Column 6. I find that those assigned no outside option exert 0.587 standard deviations more effort compared to those assigned a guaranteed outside option. Figure 1.5 presents the distribution of this index for the no outside option (T0) and the employment guarantee (T100) groups. This figure shows that the effort distribution for those guaranteed outside employment is shifted to the left. The p-value associated with a Kolmogorov-Smirnov distribution test of equality is 0.005.

In sum, I find that individuals assigned high outside options exert lower levels of effort whereas those assigned poor outside options exhibit higher effort. Therefore, the poorer performance among those with poor outside options is not driven by lower effort. These results taken together are interesting, and in the Section 1.5.4, I outline potential mechanisms that may be driving these results.

---

[39] (1.179 – 0.750)*60

1.5.3 Robustness

There are a number of specification checks that can be conducted to test the robustness these findings about the effect of employment risk on performance and effort. First, I discuss additional checks related to covariate imbalance across treatment status. Second, I attempt to address issues of missing data and differential survey response.

Covariate imbalance specification checks

Although treatment was randomly assigned and covariates appear to be balanced at baseline, given the relatively small sample there may still be persistent concerns regarding omitted variable bias. Adding covariates does not substantively alter the results further suggesting that imbalance is not a serious concern (Appendix Tables B.4 and B.5). However, as more formal specification check, I construct a ratio that measures the extent to which selection on unobservables would need to exceed selection on observables to explain away the coefficient (Altonji et al., 2005; Bellows and Miguel, 2008). A larger ratio implies that the relative omitted variable bias from unobservables relative to observables is greater, and therefore estimated effects are less likely to be explained away. Appendix Table B.5 presents the ratios for each of the performance and effort indicators for which significant differences between those assigned no outside option and those guaranteed an outside option exist.

For the case of this key performance result, the required ratio is 68, which means that the selection on unobservables would have to be 68 times greater than selection based on observables controlled for. For engagement indicators, the ratios are negative which suggests that the omitted variable bias results in an underestimate of the treatment effect rather than an overestimate. Similarly, the effort indicators suggest that selection on unobservables would have to be much larger than the selection based on observables, by ratios ranging from 7 to 9.7.

Missing administrative data and differential non-response in survey data

Though they come from administrative data, there are missing values for some of the performance and engagement indicators. For example, a subset of test score observations are missing (5.2 percent). These data are missing for various reasons including misplaced test papers; illegible or incorrect employment IDs on submitted tests; and partial training attendance that resulted in some individuals not taking all tests.[40, 41] Given that the participation rates in training are not differential across treatment groups, we would not expect the missing data to affect the results. A distinct missing data concern, as noted in Section 1.3.1 and presented in Table 1.1 Panel B, is the differential survey data completion rates. Differential completion rates by treatment group may bias the observed results where survey data were used. Conducting robustness checks in this case is particularly important.

I use the same strategies to address the concerns arising from both missing administrative data and differential non-response in the survey data. First, I present weighted results (Fitzgerald, Gottschalk and Moffitt, 1998). To do this, I first predict the probability of non-completion. Using these predicted probabilities, I construct propensity score weights for each individual. I then rerun the regressions using these weights.

[40] Recall that participation rates were not 100 percent across all training days (Table 1.1).
[41] One potential behavioral response in this setting is that job trainees assigned poor outside options reduced their participation in training, instead opting to increase external job search effort. Recall that individuals were paid for participation during the training, at a wage that is relatively competitive in this environment. While there is some evidence supporting lower attendance of individuals in the 0 percent treatment group relative to the 100 percent treatment group, the difference is neither large (4 percentage points) nor statistically significant (although the p-value is 0.147). Job search among those who attended training would have been difficult. Participants spent approximately 8 hours per day in training, and report another 1.6 hours in transit, and 6.8 hours sleeping (on average). Moreover, the job training period was conducted over a relatively short time frame, and delaying job search by three days would not be seen to be costly.

Second, I present conservative bounded results in which I implement min-max bounds (Horowitz and Manski, 1998). I impute the maximum test score for all treatment groups except for the 100 percent treatment group, for which I impute the minimum test score. In a second regression, I impute the minimum test score for all treatment groups except the 100 percent treatment group, for which I impute the maximum test score. Lastly, I restrict the sample to the 0 and 100 percent treatment groups and estimate Lee (2009) bounds on the average treatment effect of the 100 percent group relative to the 0 percent treatment group. As discussed below, these robustness checks suggest that missing data do not pose serious concerns for interpreting results for the performance and effort indicators.

Performance indicators: Appendix Table B.6 presents these results. Columns 1, 4, 7, and 10 present the weighted regressions. Columns 2, 5, 8, and 11 present conservative minimum bounds, and Columns 3, 6, 9, and 12 present conservative maximum bounds. Appendix Table B.7 Panel A presents the Lee bounds.

Overall, my three specification checks have similar findings to the main results for test scores. Point estimates from weighted results for test scores are very similar to the main results. Using the conservative min-max bounds, the differential testing performance between individuals receiving a 0 and a 100 percent chance of an alternative job is no longer statistically significant. However, the differential effect remains positive, albeit considerably smaller (Appendix Table B.6, Column 2). The Lee bounds for the average test performance results are presented in Table 1.8. In this case, I restrict the analysis to only the 0 and 100 percent treatment groups and estimate a lower bound of the performance improvement of the T100 group (compared to the T0 group) at 0.346 standard deviations and the upper bound at 0.492 standard deviations.

The results for engagement indicators are similarly robust. As for performance, the weighted results for engagement indicators are similar to the main results. Using the conservative bounding approach does not affect the direction of the coefficients, although the magnitude of the differences is muted. In addition, for the number of *good* contributions, the difference between T0 and T100 is no longer statistically significant at the 10 percent level (p=0.249). The Lee bounds for the key engagement variable, the number of *good* contributions, comparing the 0 and 100 percent treatment groups are 0.413 and 0.509.

Effort indicators: Appendix Table B.8 presents the robustness checks for the effort indicators. Columns 1, 4, and, 7 present the weighted regressions. Columns 2, 5, and 8 present minimum bounds, and Columns 3, 6, and 9 present maximum bounds. Lee bounds are presented in Appendix Table B.7. In all cases, including the time use indicators, the results discussed are robust to these rigorous specification checks. Even using the most conservative bounds for the time use results, the difference between the amount of time spent by T0 and T100 remains quantitatively large and statistically significant at the five percent level. Those assigned the job guarantee (T100) spend 19 minutes less studying the training materials, and 41 minutes more watching television or listening to the radio relative to those assigned no outside option. Particularly important in the case of the time use data, the Lee bounds show that those assigned a guaranteed outside option study the training materials less and watch more television. The upper and lower bounds are statistically significant and consistently show large differences between those assigned the job guarantee (T100) and those assigned no additional probability of outside employment (T0).

In sum, the results are generally robust to a number of alternative specifications and bounding exercises. I consistently find that performance is highest among those with guaranteed outside options and lowest among those assigned no outside option. However, effort is highest among those assigned no outside option and lowest among those with the guaranteed outside options.

1.5.4 Potential Mechanisms

The results presented thus far examine the reduced form impact of employment risk on trainees' performance and effort during the recruitment process. Contrary to the predictions of a standard economic model where performance is a decreasing function of the value of the outside option, here performance is highest among those with the best outside option (those facing no employment risk) and it is not driven by effort. In this section, I explore the alternative mechanisms that are potentially driving these results.

Essentially, the variation in outside options generated by the experiment changes the incentive to perform. Absent the outside options, performance is rewarded with employment for individuals who reach the threshold. Standard economic theory predicts that reducing the incentive to perform (by offering the outside options) should lead to decreased effort. Economic models typically assume that performance is monotonically increasing in effort, so reducing the incentive to perform should also reduce performance. Prendergast (1999) reviews the literature, which largely finds that incentives have the intended effect on the incentivized outcome, particularly in the case of simple tasks. This review touches on some cases in which incentives fail to lead to the intended outcome, and Kamenica (2012) reviews the more recent literature focusing on the empirical evidence in which incentives have anomalous effects. A number of behavioral theories have been put forth to explain these anomalous incentive effects. In this section, I discuss behavioral theories relevant to my findings and try to rule out some of the competing explanations that might be driving my results.

1.5.4.1 Incentives and stress: potential for choking under pressure

In addition to reducing the incentive to perform by increasing the value of the outside option, reduced employment risk is also likely to make the recruitment process less stressful to job seekers. The combination of reduced incentives and reduced stress leads to ambiguous predictions for performance when outside options improve.

Incentive effect

Intuitively, an improvement in an individual's outside option reduces the marginal benefit of any particular employment opportunity. Therefore, the optimal level of effort should decline as outside options improve, assuming that the cost of effort is not zero. If performance is increasing in effort, then as outside options improve, performance will decline.

In the recruitment setting I study, assume that p is the probability of being hired in the current recruitment process, and w is the wage associated with the recruiter's job. The probability of being hired is assumed to be a positive and monotonically increasing function of performance (a realistic assumption for this recruiter's hiring process); it follows that performance and employment are also positive and monotonically increasing functions of effort. Also, $1-p$ is the probability of not being successful in the recruitment process, and b is the expected value of the individual's outside option (i.e. his probability of outside employment multiplied by the expected wage of outside employment). Finally, assume that effort is costly. Therefore an individual selects effort level e^* to maximize:

$$Max_e U = p.w + (1-p)b - c(e)$$

subject to:

$$p = f(e), \qquad f'(e) > 0 \; and \; f''(e) \leq 0$$

$$c(e) \geq 0, \qquad c'(e) > 0 \; and \; c''(e) \geq 0$$

As expected, performance and employment are both rising in effort, while effort is declining in the job seeker's outside option.

Stress effect

A second key channel through which employment risk may affect performance is through its impact on stress. Extensive literatures in both psychology and public health show that unemployment is stressful, as is perceived job insecurity (Feather, 1990; de Witte 1999, 2005; Burgard et al., 2009). Therefore, it is reasonable to assume that stress is a decreasing function of an individual's outside options, i.e. *s= s(b)*, and $s' < 0$. Therefore, reducing the risk of unemployment should reduce stress.

The Yerkes-Dodson law (1908) maps the relationship between stress and performance, and performance has been shown to be inverse u-shaped in stress. As stress increases, performance improves up to a bliss point beyond which performance declines as stress continues to increase (Figure 1.6.a). Incorporating this prediction means that performance is a function of not only effort but also stress, i.e. *p = f(e,s)*. Also, f'_s varies by *s*.

As outside options improve (*b* increases), stress (*s*) should decline, but the impact on performance is ambiguous. Therefore, the stress effect induced by reduced risk should either always be positive or negative; or else it should first increase and the decrease across the risk distribution depending on the underlying values of s.

Resulting predictions combining incentive and stress effects

Because risk affects performance through both the incentive and stress channels, the predicted relationship between risk and performance is ambiguous. The sign of the net effect of employment risk on performance depends on the relative size of the incentive and stress effects, and on whether the level of risk puts the individual in the increasing or decreasing portion of the Yerkes-Dodson curve. Effort is unambiguously decreasing as the value of the outside option increases. Thus, there are three possibilities when employment risk declines:

- *Performance and effort decline:*

  If the stress effect leads to a performance decline, then the stress and incentive effects work in the same direction and effort and performance should decline. Alternatively, if the stress effect leads to a performance improvement but this is smaller in magnitude than the negative incentive effect on effort, then effort and performance will both decline.

- *No change to performance, but effort declines:*

  If the stress effect leads to a performance improvement that exactly offsets the negative incentive effect on effort, then we will observe no net effect of employment risk on performance, for a lower level of effort.

- *Improved performance, but effort declines:*

  If the stress effect leads to a performance improvement that exceeds the magnitude of the incentive effect's reduction in effort, then we will observe effort declining and performance improving.

The results from my experiment fall into the third case: I find performance improving and effort declining as outside options improve. The remaining challenge is to determine whether improved performance does, indeed, operate through the stress effect or whether some other mechanism is responsible for the change in performance. We are interested in explaining the mechanism behind the performance result, and the findings about effort can help distinguish between different possible mechanisms though not provide a conclusive test. The effort results are consistent with the presented stress mechanism but will also help to disentangle other potential mechanisms.

Ideally, to determine whether the *stress effect* really is the driver of the observed performance effects, biomarker data collection (e.g. cortisol) would have been optimal. Unfortunately, due to budgetary and logistical restrictions, this was not possible. However, in a pilot that I conducted in a similar setting, I collect four heart rate readings, at the same time of day on four different days. Two of these were taken on training days that occurred before job probabilities had been announced and the other two were taken on training days after the announcement. The repeated measurement is an attempt to reduce measurement error associated with collecting heart rates. I compare the average of the post-announcement heart rates to the average of the pre-announcement heart rates for individuals assigned job guarantees compared to those given very low chances of outside job options. Individuals assigned a guaranteed outside option experienced a 6.4 point greater decline in their heart rate (se=3.25) compared to those assigned a 1 percent outside option (in the pilot, the "no outside option" did not exist). This provides further support of a reduction in stress driven by the assigned job guarantee.

While biomarker data collection would yield insight into the presence of a biological stress response, it would not address outstanding questions regarding how stress acts to inhibit performance. Psychological research has identified many factors that contribute to sub-optimal performance, including the mere presence of an audience, public speaking, and public announcements about performance (Baumeister and Showers 1986; and Beilock, 2010). The psychological literature moves beyond identifying factors that affect performance in this way and examines precise mechanisms related to how effects on working memory lead to sub-optimal performance. In my setting, it could be that job seekers assigned the low outside option over-think their performance such that paying too much attention actually becomes counterproductive (Beilock et al, 2002). Another possibility is that individuals assigned no outside option experience an increase in distracting thoughts and worries related to their likely continued unemployment, which prevents them from focusing on the important information (Hayes, Hirsh and Matthews, 2008). This study is not designed to determine the precise mechanism through which the stress may operate to impair performance.

While my results are consistent with the theory that a reduction in stress leads to increased performance, there are a number of alternative behavioral theories that might explain the performance results. Not all of these alternatives speak to my findings on effort. These possibilities include gift exchange; stereotype-threat; the nutritional efficiency-wage hypothesis; and alternative psychological considerations. I discuss each of these in turn.

1.5.4.2 Gift exchange

A model of reciprocity provides one alternative explanation. The gift exchange hypothesis presented in seminal work by Akerlof (1982) and built upon by Akerlof and Yellen (1988 and 1990) relies on the key assumption that there is a positive relationship between wages and worker effort. This relationship explains higher than market-rate clearing prices, wherein workers reciprocate higher wages with more effort. There is substantive lab experimental evidence in support of the gift exchange model. Fehr, Kirchsteiger and Riedl (1993) provide some of the first evidence, and Fehr and Gaechter (2000) provide a survey of the reciprocity literature more generally. Recently, Gneezy and List (2006) tested the gift exchange model in the field and find only short term evidence in support the gift exchange model. They find that offering workers higher wages led to increased effort only in the first couple of hours, after which positive reciprocity was not observed.

In my setting, job trainees may feel rewarded by the recruiter when allocated a high outside option and may reciprocate by exerting more effort that, in turn, increases performance. However, although a gift exchange hypothesis yields similar performance predictions, in order for gift exchange to be the key driving mechanism, effort indicators should increase as outside options increase. I find the opposite results for effort indicators. Therefore, the higher observed performance among those assigned high outside options cannot be explained by the gift exchange mechanism.

1.5.4.3 Efficiency wage hypothesis

Another framework that would yield similar predictions for the performance results is the efficiency wage hypothesis. This hypothesis has been extensively researched (Liebenstien, 1957 and 1958; Stiglitz, 1976; Deolalikar, 1988). Improved nutritional intake improves both physical and mental well-being, which translates into increased productivity.

In the experimental setting, individuals who were guaranteed an alternative job may have been able to borrow against this guarantee and improve their nutritional intake.

The results in this paper may be attributable to better nutrition over this short period. While a comprehensive caloric intake daily roster was not administered, I did collect information on daily expenditures on food. This includes expenses on food consumed at home and away from home. Table 1.6 Columns 1 and 2 present these results. I find that food expenditures are relatively consistent across the treatment groups. I do not observe statistically significant differences in expenditures between the 0 percent and 100 percent treatment groups. When accounting for the differential survey non-response (Appendix Tables B.10) using weighted and conservative bounds, I still cannot reject that the 0 percent and 100 percent treatment groups spent equal amounts on food.[42]

Given these findings, it is unlikely that nutritional intake change is the key driver for the results observed.

1.5.4.4 Stereotype threat

The final potential explanation I consider has its origins in psychology. Steele (1997) defines stereotype threat as "the event of a negative stereotype about a group to which one belongs becoming self-relevant, usually as a plausible interpretation for something one is doing, for an experience one is having, or for a situation one is in, that has relevance to one's self-definition." A substantive literature addresses stereotype threat and test performance (Spencer et al., 1997, Maas and Cadinu, 2003; Inzlicht and Ben Zeev, 2000; Steele and Aronson, 1995).

In my setting, job trainees may perceive their outside option as a signal of their ability. Although assignment does not reveal information regarding an individual's ability or performance relative to the other participants, job trainees may still believe that

[42] There is one exception. However, the exception suggests that those assigned no outside option spend more on food as compared to those assigned the employment guarantee. This exception also works against the possibility that the observed results are driven by increased nutritional intake.

assignment is correlated with their ability.[43] In this case, performance of individuals could be driven by self-fulfilling perceptions of their own ability. This hypothesis predicts that individuals assigned low outside options are likely to perform worse, consistent with my findings.

To test whether this is the mechanism driving the performance results, I examine the extent to which job trainees updated their beliefs about getting the recruiter's job by treatment status. Respondents were asked "What percentage chance do you think you have of getting one of the available positions for the recruiter's project?" with the following options: No chance of getting a job; Less than 25 percent; Between 25 and 50 percent; 50 percent; Between 50 and 75 percent; Between 75 and 99 percent; and Certain about employment with recruiter. To create a measure of the likelihood of employment, I assign the mid-point to categories that are brackets to construct a continuous variable. For this outcome, I do not observe statistically significant differences among most groups, except for those in the 75 percent treatment group, who do report significantly higher expectations about their chances of getting a recruiter's job compared to all other groups including the guaranteed outside option group (Table 1.6, Column 4).

I also examine how the distribution of perceptions among the different treatment groups compares across time. Using Kolmogorov-Smirnov distribution tests of equality, I find that the distribution of perceptions using this measure are not different when comparing those assigned no outside option and those assigned a guaranteed outside option. In fact, with the exception of the distribution of the 50 percent probability group, none of the pairwise distribution comparisons between the various treatment groups are

[43] From open-ended questions on the survey, it is evident that respondents understood that the assignment of the outside options was not correlated with ability. For example, "The probability criteria are dependent on the chance and not merit of a person in terms of experience and qualification."; or "It is about chances."; or "Simply it's about luck".

statistically significant. Given the large number of pairwise comparisons, caution should be exercised in interpreting this result.

These results do not, however, suggest that individuals were not updating their beliefs as they underwent the recruitment process, just that individuals did not update their beliefs differentially by treatment status.

However, again the starkest key performance differential is observed between individuals in the 0 percent treatment group and those in the 100 percent treatment group, and this does not seem to be driven by stereotype threat, as there are not large differences in these two groups' perceptions of their chance of being of being hired for the recruiter's job.

Neither gift exchange mechanisms, nor efficiency wages, nor stereotype threat can explain the full pattern of my results. Instead, it seems that the most plausible mechanism driving the performance and effort results is a framework in which the varied outside options reduce effort, and simultaneously reduce stress, enabling a higher return to effort.

1.5.5 Welfare Implications: Employment

It is important to assess the welfare implications of the observed performance response to employment risk. To do this, I examine employment outcomes.[44] As discussed in Appendix A, while the performance indicators do a relatively good job of

[44] There are a number of on-the-job performance measures that can be constructed, i.e. performance on-the-job when successfully hired and working for the recruiter. On-the-job performance was not measured during the alternative jobs, however the long term impacts of being assigned an alternative job on future employment and wages are presented in Godlonton (2013). Recall that job trainees were hired as interviewers for a health survey. Therefore, to measure on-the-job performance, one can use survey data from the health survey. For example, one can measure the number of skip rules incorrectly followed and the number of inconsistencies by interviewer. For these indicators, there is little difference by treatment group. A summary measure of performance is whether the recruiter offers an individual a renewed short term employment contract. The recruiter had three waves of contract renewals. In general, the likelihood that individuals in the 75 and 100 percent groups are hired in each subsequent round is about twice the employment rates of the other groups. Differences are often not statistically significant due to limited power.

predicting performance, there is still a large unobserved component determining employment outcomes.

Figure 1.7 depicts the share of job trainees hired by the recruiter by treatment group. About 25 percent of trainees in the 75 and 100 percent groups were offered employment by the recruiter.[45] Thirteen percent of individuals who received no chance of an alternative job were hired by the recruiter. Those individuals who received a 50 percent chance of an alternative job were the least likely of all treatment groups to be hired by the recruiter – only 11 percent of these participants were hired by the recruiter, making them half as likely to be hired relative to those who knew they had high chances of alternative employment.

Table 1.7 Panel A presents the OLS results for employment as depicted in Figure 1.7. Table 1.7 Panel B presents the probit results. The marginal effects reported are the partial derivatives evaluated at the mean of the covariates. Given the performance indicator results, it seems reasonable to use the 100 percent treatment group as the omitted category. The results are similar in the full sample as compared to a restricted sample that consists only of trainees who attended training every day. Individuals in the 0, 5, and 50 percent chance of alternative work treatment groups are less likely to be employed by the recruiter by between nine and 11 percentage points. These impacts are statistically significant and are large in magnitude, as they translate to a 50 percent lower chance of being hired than those in the 100 percent treatment group.

Two other results are worth noting. First, individuals assigned to the 75 percent treatment group are no less likely than those in the 100 percent group to be hired by the

[45] Note that only one participant who was offered a position by the recruiter opted not to take the job. As such the offer of a job and the record of who got hired are approximately the same.

recruiter. Recall that, on average, this group did not perform well on the written tests, but there is considerable heterogeneity in their performance both by mental health status and ability (see Section 1.5.6). Second, there is suggestive evidence that the individuals in the 1 percent treatment group are more likely to be recruited than those in the 0 percent treatment group. Although there is insufficient power in the current sample to determine this, it is interesting to note that a small change potentially has large impacts.

1.5.5.1 Potential confounders for employment results

One threat to the interpretation of the employment results is the potential of strategic behavior by the recruiter in his hiring decisions in response to treatment assignments. However, the recruitment team had no knowledge of the specific alternative job probability assigned to each participant. The only way the recruitment staff would know of a trainee's alternative job probability is if that participant directly informed a recruiter. Anecdotally there are no reports of this occurring. Even if it did, one would expect that it would bias the results in favor of higher employment rates for those assigned lower alternative job probabilities. Given that I observe lower employment rates in this group, if such strategic behavior had been present, my results are a downward biased estimate.

Another concern is that, assuming that the recruiter did learn of a trainee's alternative job probability, he may have (incorrectly) inferred that a high probability of an alternative job implied something about the ability of the trainee. The recruiter has implemented randomized controlled trials for a number of years within Malawi and understands the concept of random assignment. Moreover, the recruiter provided the ability scores in order that random assignment of treatments was stratified across baseline. As such, strategic behavior from the recruiter's perspective based on the random assignment is unlikely whether or not the trainees tried to lobby in any particular way.

1.5.6 Heterogeneity of Performance and Employment Differences

Thus far, we have established that reduced employment uncertainty improves welfare by increasing the chance of being hired, which is unsurprising given that reduced uncertainty also leads to higher performance. I turn now to examining heterogeneous responses to employment risk. Understanding the heterogeneity in the effect of uncertainty on performance and employment may have important policy considerations or distributional implications.

Heterogeneity may arise for a number of reasons. Job seekers likely face different cost of effort functions. For example, the cost of effort may be dependent on ability, resulting in differential effort responses to reductions in risk.

Also, research in psychology finds that individuals differ in their responses to stress (Hobfoll, 2004). Specifically, there is likely to be heterogeneity in the stress effect induced by the reduction in employment risk for two different reasons. First, there is variation in baseline stress, and previous literature shows that the response to stress is non-linear. For example, compare individuals at t'_0 and t''_0 on the Yerkes-Dodson curve illustrated in Figure 1.6.a. A reduction in stress of amount *s* results in differential changes in performance for these individuals. Second, even among individuals with the same baseline stress level, extensive research shows that individuals differ in their ability to cope with stress (Ditzen et al., 2008; Fiocco, Joober and Lupien, 2007). Therefore, the same change in employment risk may yield differential stress reductions across individuals. For example, consider two different individuals at t'_0 in Figure 1.6.b. For one individual, the employment guarantee may reduce stress by *s*. For a second individual, it may reduce stress by s', resulting in different implications for performance.

Clearly, there are multiple sources of potential heterogeneity. While I cannot, in this setting, separately measure the extent to which there is heterogeneity in the stress and the incentive effects, I can show how the reduced form relationship between employment risk, performance, and employment differs for different types of job seekers. I focus on how performance and employment differ by baseline mental health status and ability. It is important to highlight that my power to detect differences is limited and results should be interpreted with caution. I first adopt a simple approach by classifying individuals as exhibiting high/low mental health and, separately, high/low baseline ability. I then plot the average performance and employment by treatment group for these groups. Then I present regression results from the following regression:

$$Y_i = \beta_1 T0_i + \beta_2 T1_i + \beta_3 T5_i + \beta_4 T50_i + \beta_5 T75_i + \beta_6 T100_i + \beta_7 T0_i * HET_i + \beta_8 T1_i * HET_i + \beta_9 T5_i * HET_i + \beta_{10} T50_i * HET_i + \beta_{11} T75_i * HET_i + \beta_{12} T100_i * HET_i + X_i{}'\beta + \varepsilon_i \qquad (3)$$

where: all treatment dummy indicators are interacted with a job seeker attribute (Het). In one set of results, the *Het* variable is a measure of baseline mental health. I present specifications using a binary measure of good mental health as well as a standardized continuous mental health score. Similarly, in the second set of results, *Het* measures baseline ability using either an indicator of high ability or a standardized measure of baseline ability. This approach assumes that any risk-performance differences are linear in ability or mental health. To explore whether this is a problematic assumption, I present non-parametric regressions of the difference between individuals assigned a guaranteed outside option and no outside option across the mental health and baseline ability distribution. The assumption seems reasonable for the mental health results but not for the ability results. To allow more flexibility for the risk-performance relationship to vary across ability, I split the sample into three ability groups – low ability, medium ability, and high ability and present the treatment group averages for each group.

Mental health

I examine variation by baseline mental health status as there is extensive research showing that long term and short term stressors have different impacts and interact in important ways. Individuals with better mental health are better able to cope when faced with employment uncertainty and, in this case, may incur a smaller benefit from the stress reduction of the employment guarantee. However, given that mental health and stress are highly correlated, individuals with better mental health may exhibit lower baseline stress levels and therefore may benefit more from the stress reduction due to the concavity of the Yerkes-Dodson curve. Therefore, it is ambiguous how effects may differ across groups that differ by mental health status.

To measure mental health status, I use the SF-36 instrument that maps into eight health indicators. Four pertain to mental health and can be used to construct a composite mental health summary measure; four pertain to physical health (Ware and Sherborne, 1992; Ware, Kosinski, and Keller, 1994 and 1995). This instrument has been widely used worldwide and validated in other African countries (Wagner et al. 1999; Wyss et al. 1999). Because the mental health composite measure has been shown to predict mental health problems as well as or better than the individual mental health indices (Ware et al., 1995), I use the composite index. The mental health index takes on values from zero to

100 where a higher number represents better mental health. In the sample, I observe a range of 39 to 81. I also construct a binary indicator of "good" mental health, which is equal to one if the individual scores above the mean mental health score in the sample.

Figure 1.8 presents a bar graph of the average performance by mental health type (good or poor) and treatment group. First, on average, individuals exhibiting poorer mental health perform worse than those with better mental health, regardless of their assigned outside option.[46] Second, among those who exhibit good mental health, performance increases in the probability of the outside option. However, for those with poorer mental health status, the relationship between performance and risk is non-monotonic.

Table 1.8 presents regression results controlling for covariates and shows that the gap between the performance of individuals with better mental health (compared to those with poorer mental health) is weakly larger when assigned the guaranteed outside option compared to no outside option. To illustrate this, consider the following test: $\beta_1 + \beta_7 = \beta_6 + \beta_{12}$. The p-value associated with this test is 0.099. However, this result is not robust to using a binary indicator of "good" mental health (p=0.377).

Figure 1.9 presents the fan regression of the difference between the performance among those assigned a guaranteed outside option and those assigned no outside option, across the mental health distribution. This graph is trimmed to achieve common support on either end of the mental health distribution among the two treatment conditions. However, this graph merely serves to illustrate that there do not seem to be significant

[46] This relationship is not driven by correlation between ability and mental health. Although Figure 1.8 does not control for any covariates, Table 1.8 does, and this relationship persists.

differences in the performance differential when eliminating risk across the mental health distribution. Therefore, the imposed linearity in Table 1.6 seems appropriate.

While test performance was the most important predictor of employment, it is useful to examine heterogeneity in employment outcomes by mental health status. Figure 1.10 presents the average employment rate by mental health type and treatment group. Across most of the treatment conditions, the differences by mental health status in the fraction employed are mostly small and statistically insignificant. Only in the case of the group assigned the 75 percent chance of outside employment do we observe marginally significant (although quantitatively large) differences. In sum, although there appear to be large differences in test scores by mental health status, these do not translate into differential employment outcomes in this sample. It is beyond the scope of the current experiment to determine how individuals with lower mental health are able to compensate for their poor test performances and achieve equal probability of employment by the recruiter, but this is an important avenue for future research.

Baseline Ability

Next, I turn to examining heterogeneity by baseline ability. Given that I stratified treatment assignment by ability and prior work experience with this recruiter, these are two obvious dimensions to explore heterogeneity of the performance and employment impacts. My power to detect differences by familiarity with this specific recruiter is limited, since only 10 percent of the sample (26 individuals) had prior work experience with the recruiter.

Figure 1.11 presents the test results by ability type and treatment group graphically.[47] I find that among low ability types, reducing the risk of unemployment generally increases performance. The extremely poor performance by those assigned a 75 percent chance of an alternative job is hard to explain and may simply be an artifact of the data. The relationship between risk and performance is non-linear among the high ability types.

Table 1.8 presents the average performance by treatment group from the regressions, controlling for other covariates, and the results support the interpretation of the graphs. As with the mental health results, these specifications assume a linear relationship across the ability distribution. Unlike with the mental health results, the fan regression (Figure 1.12) plotting the difference in performance for those assigned the guarantee versus no outside option across the ability distribution suggests that this linear assumption is not consistent with the data. This figure suggests that the largest differences are incurred by the job seekers in the middle of the ability distribution. I therefore split the sample into three groups – low ability, medium, and high ability – and present the performance and employment differences for these three types.

Figure 1.13 presents the average performance by ability type using these three categories. I find the increasing trend in performance for the high and low ability types (with the exception of the 75 percent treatment group for the low ability types). I also find that the relationship is non-linear for those in the middle of the ability distribution.

Figure 1.14 presents the average employment outcomes using this three ability type classification. Here we observe large differences in employment across the different

[47] For the purposes of examining stress or choking under pressure, this selection test is not a good measure of ability as it too was a high stakes test, and individuals prone to choking under pressure may have performed sub-optimally on this test.

treatment groups. High ability types are the most likely to be hired, and their employment rates are least affected by the varied employment risk. The risk appears to affect individuals in the middle of the distribution most dramatically: this group benefits the most from the eliminated employment risk (the guaranteed outside option). While reduced risk also benefits low ability types, those differences are marginal.

These results suggest that individuals at either end of the distribution are the least affected by changing employment risk, in terms of employment losses, and individuals in the middle of the ability distribution are the most susceptible to such risk.

1.6 Conclusion

I find that job seeker performance during recruitment is highest and effort is lowest among those assigned the best outside employment options, while the converse is true for those assigned the worst outside options. The latter group of job trainees both perform better on tests of materials taught during training and are more actively engaged in the recruitment process. However, these improvements are not driven by changes in effort, and are not linear in the probability of outside employment.

These findings are consistent with prior laboratory evidence (Ariely et al. 2009) that observed lower performance under high stake incentives. However, I observe this relationship in a real environment where the risk I study is real. The variation in risk in laboratory studies is artificial and over windfall income, but in my setting, the variation is over risk in securing real, meaningful employment equivalent to that for which subjects have chosen to apply through a competitive and arduous process. To my knowledge, no evidence in real-world settings has illustrated the link between risk, performance, and effort, and I provide the first evidence that previous findings do extend beyond the laboratory into real-world labor markets, something noted as an open question as recently as Kamenica (2012). There are many possible extensions to this research now that I have moved it to a real-world setting. My results examine performance during recruitment; how performance may be affected on-the-job is an important and interesting avenue for further research. Also, my results are obtained in a context in which cognitive performance is important. Whether such results will be observed in manual rural labor markets is also interesting, both theoretically, as it pertains to the mechanism through which uncertainty affects performance, and practically because of its policy relevance to the large fraction of adults in developing countries who do manual labor.

My paper also contributes to conceptual questions about the relationship between risk and performance. My results suggest considerable non-linearities in the relationship between performance and risk, which deserve further attention. Because realized outcomes are binary, studies conducted using secondary data typically do not observe the full distribution of uncertainty between an event occurring with probability zero and it occurring for certain. My results suggest that conclusions about the relationship between risk and performance are sensitive to the range of risk observed. Moreover, the observed relationship between risk and performance also has implications for how to model behavior under uncertainty. Typically, when we consider risk in theoretical models, it is modeled as a parameter of the utility function. My results do not reject this approach, but they do imply that we should also consider risk in production functions.

While the reduced-form effect of risk on performance is interesting in its own right and has real-world welfare implications, I also explore the mechanisms that might be driving the key results I observe. Using rich baseline and outcome data, I combine economic and psychological insights to explore potential mechanisms through which risk and uncertainty affect behavior. My findings suggest that the relationship between risk and performance is likely driven by a stress response. However, unlike laboratory evidence that directly measures stress using biomarkers such as cortisol (for example, Angelucci, et al. 2012), I was not able to measure hormonal stress in this manner. That said, my results do not seem to be driven by models of reciprocity, the efficiency-wage hypothesis, or stereotype threat. I cannot rule out that some other psychological consideration that operates similarly to stress is driving the result. Moreover, I cannot determine the precise psychological mechanism through which stress operates, i.e. distraction or over-thinking. Future research that more precisely measures stress would be a natural extension of this work.

Finally, while my paper is most closely tied to the laboratory experiments about the effect of risk and stress on performance, my study also speaks to the growing literature about the effect of high-stakes testing. In my study, performance under high stakes (low probability of an outside job) is worse than performance under low stakes (high probability or guarantee of an outside job). There is a growing body of literature demonstrating heterogeneous responses to high stakes vs low stakes settings. For example, Ors et al (forthcoming) show large gender disparities in low stakes vs high stakes testing situations. In low stake testing environment females outperform males; however, the same females perform sub-optimally and, on average, worse than the males on a high stakes entrance exam to an elite university. I have limited power to detect differences by ability and mental health status. My results suggest that performance differences differ by mental health status, but these differences do not translate into differential employment outcomes. However, differences by ability are important for employment outcomes; in particular, individuals in the middle of the distribution are the most affected by the reduction in risk. Understanding which types of individuals are most susceptible to risk-related performance declines could have substantive policy implications for job training or recruiting processes and deserves further attention in future research.

1.7 References

Akerlof, G.A. (1982): “Labor Contracts as Partial Gift Exchange”, The Quarterly Journal of Economics, Vol. 97, No. 4, pp. 543 – 569.

Akerlof, G.A. and Yellen, J. (1988): “Fairness and Unemployment”, American Economic Review, Papers and Proceedings, Vol. 83, pp. 44 – 49.

Akerlof, G.A. and Yellen, J. (1990): “The Fair Wage-Effort Hypothesis and Unemployment”, Quarterly Journal of Economics, Vol. 105, pp. 255 – 283.

Altonji, J.G., Elder, T.E. and Taber, C.R. (2005): “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools.” Journal of Political Economy 113(1): 151–184.

Angelucci, M., Breman, A., Cordova, K and Ordonez, L. (2012): “Mental Overload and Entrepeneurship”, University of Michigan working paper

Ariely, D., U. Gneezy, G. Loewenstein and N. Mazar (2009). “Large Stakes and Big Mistakes,”The Review of Economic Studies, 76, 451-469.

Apesteguia, J. and Palacios-Huerta,I. (2010) “Psychological Pressure in Competitive Environment: Evidence from a Randomized Natural Experiment”, American Economic Review, Vol. 100(5) pp. 2548 – 2564.

Band, S.R. and Manuele C.A. (1987), Stress and Police Officer Performance: An Examination of Effective Coping Behavior, Police Studies Vol. 10(3) pp. 122 – 131.

Baumeister, R.F. “Choking under Pressure: Self-consciousness and Paradoxical Effects of Incentives on Skillful Performance.” Journal of Personality and Social Psychology, 46(3), March, 610-620.

Baumeister, R.F., and Showers, C.J. (1986), “A review of paradoxical performance effects: Choking under pressure in sports and mental tests,” European Journal of Social Psychology, XVI (October-December), 361-383.

Beaman, L. and Magruder, J. (forthcoming), “Who gets the job referral? Evidence from a social networks experiment,” American Economic Review.

Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002): “When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. Journal of Experimental Psychology: Applied, 8, 6–16

Beilock, S. (2010) “Choke: What the secrets of the brain reveal about getting it right when you have to,” New York: Free Press.

Bellows, J. and Miguel, E. (2009) “War and Local Collective Action in Sierra Leone,” Journal of Public Economics, Vol. 93(11-12) pp. 1144 – 1157.

Brenner, H. 1971. "Economic Changes and Heart Disease Mortality." American Journal of Public Health 65: 606–11.

------. 1975. "Trends in Alcohol Consumption and Associated Illnesses: Some Effects of Economic Changes." American Journal of Public Health 65: 1279–92.

------. 1979. "Mortality and the National Economy." The Lancet 26: 568–573.
Burgard, S., Brand, J.E. and House, J.S. (2009): “Perceived job insecurity and worker health in the United States”, Social Science and Medicine Vol. 69 pp. 777 – 785.
Burns, J., Godlonton, S. and Keswell, M. (2010) “Social Networks, Employment and Worker Discouragement: Evidence from South Africa”, Labour Economics, 17, pp. 336 – 344.
Cameron, C.A., Gelbach, J.B. and Miller, D.L. (2008) “Bootstrap-based improvements for inference with clustered errors. Review of Economics and Statistics, 90(3):414- 427.
Carroll, C. D. (1994): “How Does Future Income Affect Current Consumption?” Quarterly Journal of Economics, Vol. 109 pp. 111–148.
Chan, S. and Stevens, A.H. (2001): “Job loss and employment patterns of older workers”, Journal of Labor Economics, Vol. 19 (2) pp. 484 – 521.
Chinkhumba, J., Godlonton, S. and Thornton,R. (2012): "Demand for Medical Male Circumcision." Working Paper, University of Michigan.
Couch, K.A. (2001): “Earnings Losses and Hours Reductions of Displaced Workers in Germany.” Industrial and Labor Relations Review, 54(3): 559-572.
Dee, T. and West, W.R. (2011): “The non-cognitive returns to class size.” Educational Evaluation and Policy Analysis, 30(1), pp. 23 – 46.
Deolalikar, A.B. (1988): “Nutrition and Labor Productivity in Agriculture: Estimates for Rural South India”, Review of Economics and Statistics, Vol. 70(3).
de Witte, H. (1999): “Job insecurity and psychological well-being. Review of the literature and exploration of some unresolved issues”, European Journal of Work and Organizational Psychology, Vol. 8(2), pp. 155 – 177.
de Witte, H. (2005): “Job insecurity: Review of the International Literature on definitions, prevalence, antecedents and consequences”, SA Journal of Industrial Psychology Vol. 31(4) pp. 1 -6.
Ditzen, B., Hoppman, C. and Klumb, P. (2008): “Positive couple interactions and daily cortisol: on the stress-protecting role of intimacy”, Clinical Psychology and Psychotherapy, Vol. October 70(8) pp. 883 – 889.
Dohmen, T.J. “Do Professionals Choke Under Pressure?” Journal of Economic Behavior and Organization, 2006, doi:10.1016/j.jebo.2005.12.004.
Dooley, D., Catalano, R. and Wilson, G. (1994): "Depression and Unemployment: Panel Findings from the Epidemiologic Catchment Area Study." American Journal of Community Psychology 61 (3): pp. 745–65.
Dooley, D., Prause, J. and Ham-Rowbottom, K. (2000): "Inadequate Employment and High Depressive Symptoms: Panel analyses." International Journal of Psychology Vol. 35: 294–294.
Dynan, K.E. (1993): “How Prudent Are Consumers?” Journal of Political Economy, Vol. 101, pp. 1104–1113.

Edmark, K. (2005): "Unemployment and Crime: Is There a Connection?", The Scandinavian Journal of Economics, Vol. 107, No. 2 pp. 353 - 373.

Fafchamps, M. (2010): “Vulnerability, Risk Management and Agricultural Development,” African Journal of Agricultural Economics, Vol. 5(1).

Feather, N.T. (1990): “The Psychological Impact of Unemployment”, Springer series in social psychology. New York, NY, US: Springer-Verlag Publishing.

Fehr, E. and Gaechter, S. (2000): “Fairness and Retaliation: The Economics of Reciprocity”, Journal of Economic Perspectives, Vol. 14, pp. 159 – 181.

Fehr, E., Kirchsteiger, G. and Riedl, A. (1993): “Does Fairness Prevent Market Clearing? An Experimental Investigation”, Quarterly Journal of Economics, Vol. 108, pp. 437 – 460.

Fiocco, A. J., R. Joober, Lupien, S.J., et al. (2007). "Education Modulates Cortisol Reactivity to the Trier Social Stress Test in Middle-Aged Adults." Psychoneuroendocrinology 32(8-10): 1158-63.

Fitzgerald, J. and Gottschalk, P. and Moffitt, R. (1998): "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," Journal of Human Resources, University of Wisconsin Press, vol. 33(2), pp. 251-299.

Gneezy and List (2006): “Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets using Field Experiments”, Econometrica, Vol. 74 No.5 pp. 1365 – 1384.

Godlonton, S. (2013): “Returns to Work Experience: Evidence from Malawi,” University of Michigan Working Paper.

Green, D.A. and Riddell, W.G. (1993) “The Economic Effects of Unemployment Insurance in Canada: An Empirical Analysis of UI Disentitlement”, Journal of Labor Economics, Vol. 11, No. 1, Part 2: U.S. and Canadian Income Maintenance Programs (Jan., 1993), pp. S96-S147

Gregg, P. and Tominey, E. (2005) “The wage scar from male youth unemployment,” Labour Economics Vol.12(4), August, pp. 487–509.

Gruber, J. (1997): “The Consumption Smoothing Benefits of Unemployment Insurance,” American Economic Review, Vol. 87(1) pp. 192 – 205.

Heyneman, S.P. et al. (1981): “Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement,” Journal of Education Psychology, Vol. 73, No. 4, pp. 556-567.

Hayes, S., Hirsch, C. and Matthews, A. (2008): “Restriction of Working Memory Capacity during Worry,” Journal of Abnormal Psychology, Vol. 117(3): 712-7.

Hobfoll, S. E. (1998). Stress, culture, and community. New York: Plenum.

Horowitz, J. L., and Manski, C. F. (1998), “Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations,” Journal of Econometrics, 84, 37-58.

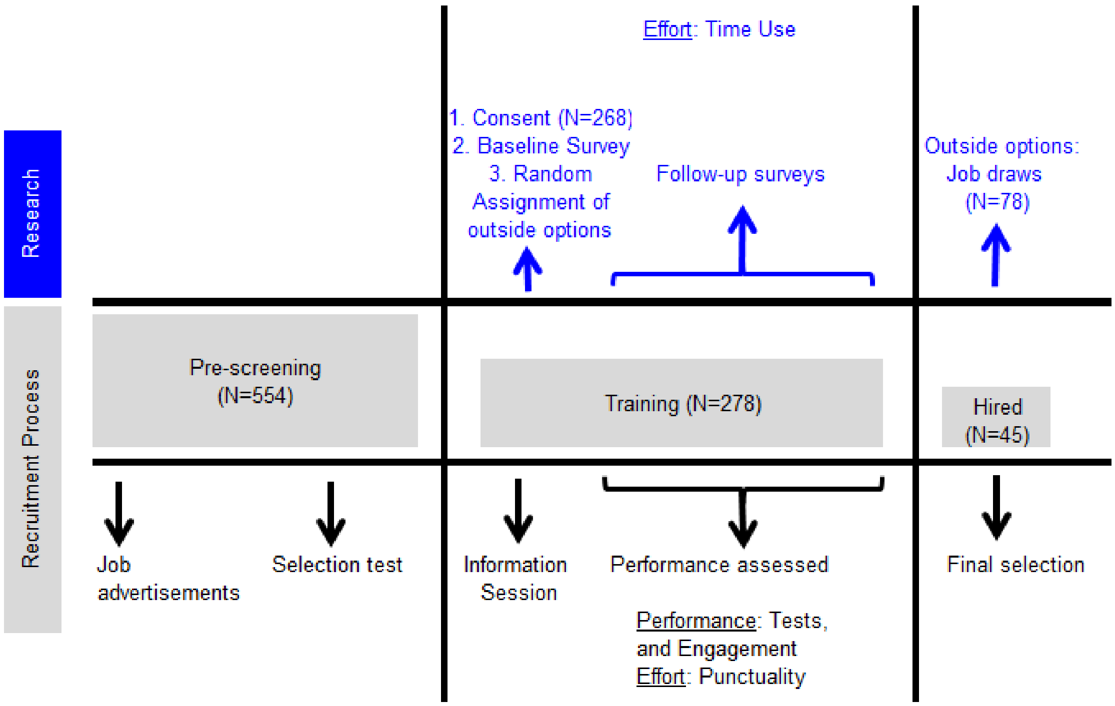
Human Development Report. (2009): “Overcoming barriers: Human mobility and

development," Published for United Nations Development Programme.
ILO. (2012): "Global Employment Trends 2012: Preventing a deeper job crisis", ILO: Geneva
Integrated Household Survey (IHS). (2010/11): "Third Integrated Household Survey Data conducted by Malawi National Statistics Office in collaboration with the World Bank.
Inzlicht, M. and Ben-Zeev,T. (2000): "A threatening intellectual environment: Why Females are susceptible to experiencing problem-solving deficits in the presence of males", Psychological Science, Vol. 11, No. 5, pp. 365 – 371.
Jamal,M. (1984) "Job stress and job performance controversy: An empirical assessment", Organizational Behavior and Human Performance, Volume 33, Issue 1, February 1984, Pages 1-21
Jensen, P. and Smith, N. (1990) "Unemployment and Marital Dissolution," Journal of Population Economics, Vol. 3, No. 3. pp. 215 – 229.
Jin, R.L. and Shah, C.P. and Svoboda, T.J. (1995) "The impact of unemployment on health: a review of the evidence," CMAJ Vol. 135(11).
Kamenica, E.(2012) "Behavioral Economics and Psychology of Incentives", Annual Review of Economics Vol. 4, pp. 427-52.
Kimball, M. and Willis, R. (2006): "Utility and Happiness", University of Michigan working paper.
Kling, J.R., Liebman, J.B. and Katz, L.F. (2007) "Experimental Analysis of Neighborhood Effects", Econometrica, Vol. 75(1) pp. 83 – 119.
Kremer, M. and Holla, A. (2008): "Improving Education in the Developing World: What Have We Learned from Randomized Evaluations?" in Kenneth J. Arrow and Timoothy F. Bresnahan, eds., Annual Review of Economics, Volume One.
Lee, D. (2009):"Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", Review of Economic Studies, 2009, 76(3), 1071-1102.
Leibenstein, H. (1957): "The Theory of Underemployment in Backward Economies." Journal of Political Economy, vol. 65.
Leibenstein, H. (1958): "Underemployment in Backward Economies: Some Additional Notes." Journal of Political Economy, vol. 66.
Linn, M.W., Sandifer, R. and Stein, S. (1985) "Effects of unemployment on mental and physical health," Am J Public Health. May; 75(5): pp. 502–506.
Lundin, A., and Hemmingsson, T. 2009. "Unemployment and Suicide." The Lancet 374: 270–1.
Lusardi, A. (1998): "On the Importance of Precautionary Saving," American Economic Review, Vol. 88 pp. 448–453
Maas, A. and Cadinu, M. (2003): "Stereotype threat: When minority members underperform", European Review of Social Psychology, Vol. 14. Pp. 243- 275.
Magruder, J. (2012): "High Unemployment Yet Few Small Firms: The Role of

Centralized Bargaining in South Africa”, American Economic Journal: Applied Economics, vol 4(3) pp 138-66.
Manning, A. and Saidi, F. (2008): “Understanding the Gender Pay Gap. What's Competition Got to Do with It?” CEP Discussion Paper.
Motowidlo, Stephan J.; Packard, John S.; Manning, Michael R. “Occupational stress: Its causes and consequences for job performance,” Journal of Applied Psychology, Vol 71(4), Nov 1986, pp. 618-629. doi: 10.1037/0021-9010.71.4.618
Murphy G.C., and Athanasou J.A. 1999. "The Effect of Unemployment on Mental Health." Journal of Occupational and Organizational Psychology 72: 83–99.
Örs, E., F. Palomino, and E. Peyrache (forthcoming). “Performance Gender-Gap: Does Competition Matter?”, Journal of Labor Economics.
Paserman, D. (2010) “Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players.” Discussion Paper, Boston University.
Raphael, S. and Winter-Ebmer, R. (2001): “Identifying the Effect of Unemployment on Crime”, Journal of Law and Economics, Vol. 44. No. 1 pp. 259 – 283.
Ruhm, C.J. (1991b): “Are workers permanently scarred by job displacements”, American Economic Review, Vol. 81 pp. 319 – 324.
Ruhm, C.J. (1994): “Advance Notice, Job Search, and Postdisplacement Earnings”, Journal of Labor Economics, 12(1), January 1994, 1-28.
Ruhm, C.J. (2000): “Are Recessions Good For Your Health?”, Quarterly Journal of Economics, 115(2), May 2000, 617-650.
Sandmo, A. (1971): “On the Theory of the Competitive Firm Under Price Uncertainty. American Economic Review, Vol. 61(1): pp. 65 - 73.
Sarason, G. and Mandler, S.B. (1952): “A study of anxiety and learning,” Journal of Abnormal and Social Psychology, Vol. 47, pp.166-73.
Segal, C. (2008): “Classroom Behavior,” Journal of Human Resources, 43(4), 2008, pp. 783-814.
Segal, C. (2012): “Working When No One is Watching: Motivation, Test Scores, and Economic Success”, University of Zurich working paper.
Segal, C. (forthcoming): “Misbehavior, Education, and Labor Market Outcomes”, The Journal of the European Economic Association.
Spencer, S., Steele, C.M, Quinn, D. (1999): “Under Stereotype threat and women’s math performance”, Journal of Experimental Social Psychology, Vol. 35 pp. 4-28.
Steele, C.M. (1997): “A threat in the air: How stereotypes shape intellectual identity and performance,” American Psychologist, Vol. 6, pp. 613 – 629.
Steele, C.M. and Aronson, J. (1995): “Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology, Vol. 29 pp. 797-811.
Stevens, A.H. (1997): “Persistent effects of job displacement: The importance of multiple

job losses”, Journal of Labor Economics, Vol. 15 pp. 165 – 188.
Stiglitz, J.E. (1976). “The efficiency wage hypothesis, surplus labour and the distribution of labour in LDCs.” Oxford Economic Papers, vol. 28.
Sullivan, J.X. (2008): “Borrowing during Unemployment: Unsecured Debt as a safety net”, Journal of Human Resources, Vol. 43 (2) pp. 383 – 412.
Sullivan, S.E. and Bhagat, R.S. (1992): “Organizational Stress, Job Satisfaction and Job Performance: Where do we go from here?”, Journal of Management Vol. 18(2) pp. 353 – 374.
Topel, R. (1990): “Specific Captial and Unemployment: Measuring the Costs and Consequences of Job loss”, Carnegie-Rochester Conference series on Public Policy, Vol 22, pp. 181 – 214.
Tryon, G.S. (1980): “The Measurement and Treatment of Test Anxiety,” Review of Educational Research, Vol. 50(2) pp. 343 – 372.
Wagner, A.K. (1999): “A Kiswahili version of the SF-36 Health Survey for use in Tanzania: translation and tests of scaling assumptions”, Quality of Life Research Vol. 8 (1-2) pp. 101- 110.
Ware JE, Kosinski M, Keller SD. SF-12®: How to Score the SF-12® Physical and Mental Health Summary Scales. Second Edition. Boston, MA: The Health Institute.
Ware JE, Kosinski M, Keller SK. (1994): “SF-36® Physical and Mental Health Summary Scales: A User's Manual” Boston, MA: The Health Institute.
Ware JE, Sherbourne CD. (1992): “The MOS 36-Item Short-Form Health Survey (SF-36®): I. conceptual framework and item selection,” Med Care; 30(6):473-83.
World Development Report: Jobs (2013). World Bank.
Wyss, K. et al. (1999): “Validation of Kiswahili version of the SF-36 Health Survey in a representative sample of an urban population in Tanzania”, Quality of Life Research Vol. 8 (1-2) pp. 111- 120.
Yerkes, R. M. and Dodson, J.D. (1908): “The relationship of strength of stimulus to rapidity of habit-formation,” Journal of Comparative Neurology of Psychology, pp. 459-482.

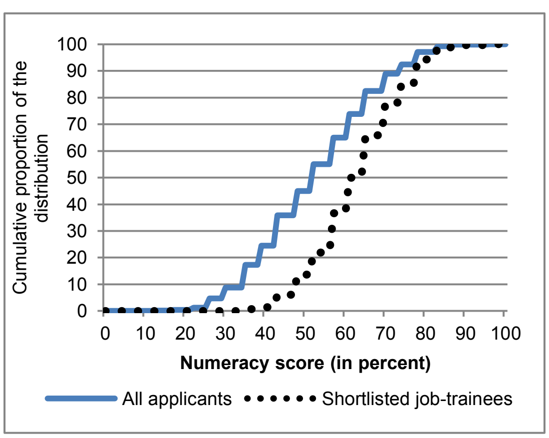
Figure 1.1: Timeline of recruitment and research activities



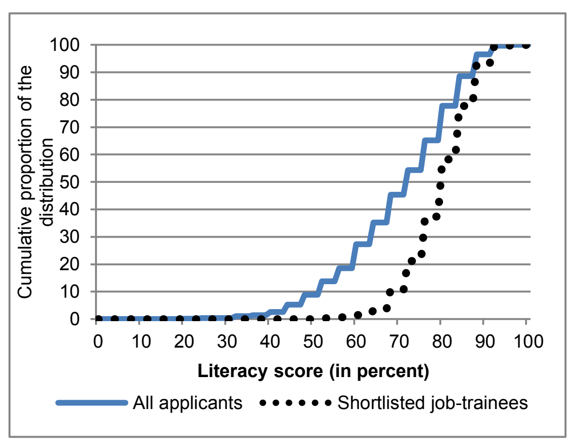
Note: Items in blue indicate research activities conducted for the purposes of this study. Items in black indicate standard recruitment activities performed by the recruiter.

Figure 1.2: Distribution of numeracy, literacy and ability scores

a: Distribution of numeracy scores



b: Distribution of literacy scores



c: Distribution of baseline ability score

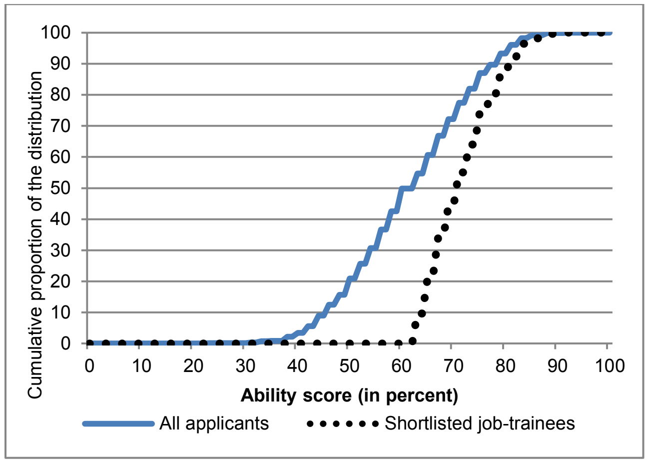
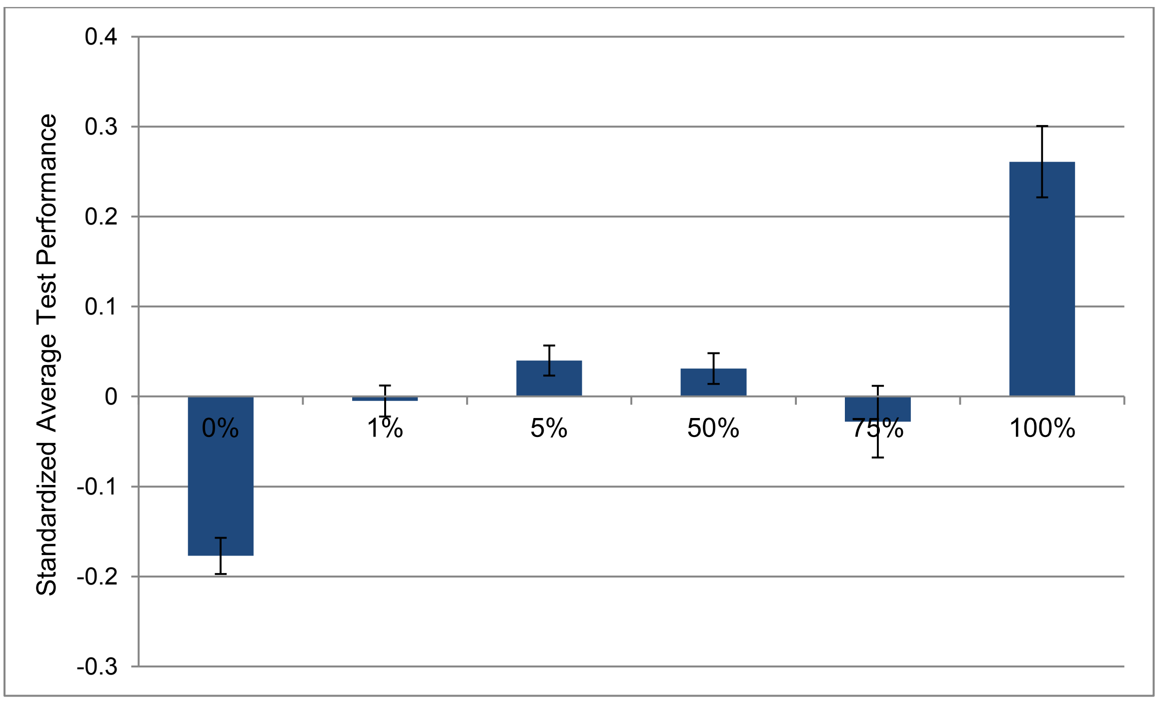


Figure 1.3: Average standardized test score by treatment group



Note: This figure presents the estimated group means controlling for covariates and stratification cell fixed effects.

Figure 1.4: Performance Index Distribution

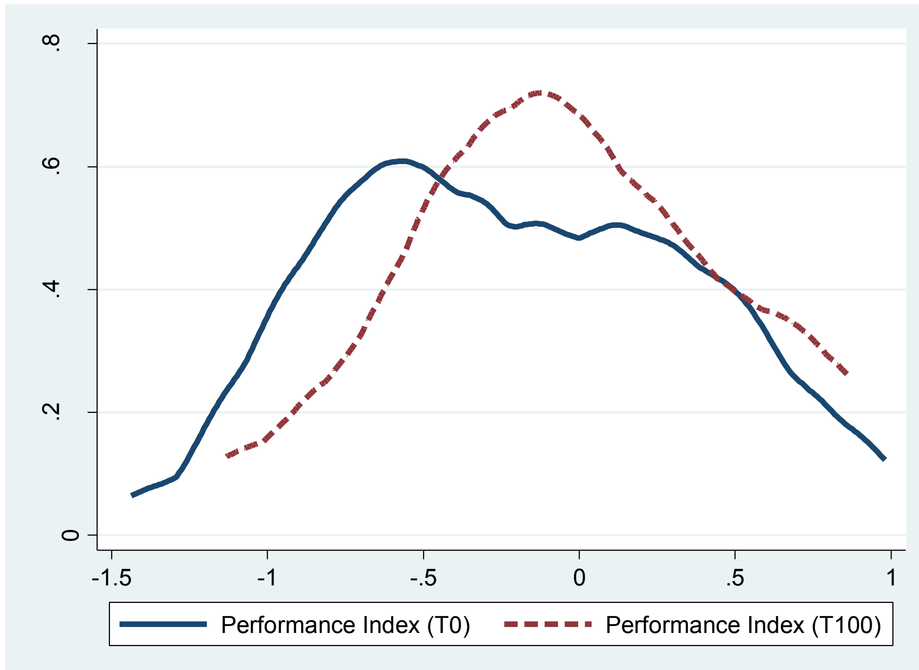


Figure 1.5: Effort Index Distribution

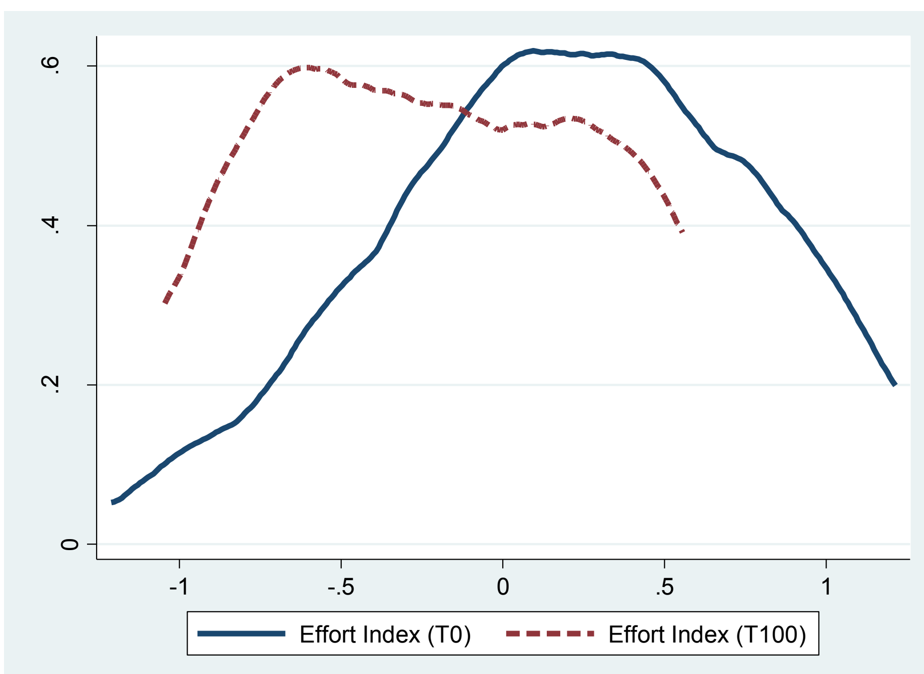


Figure 1.6: Yerkes-Dodson (1908): Relationship between Stress and Performance

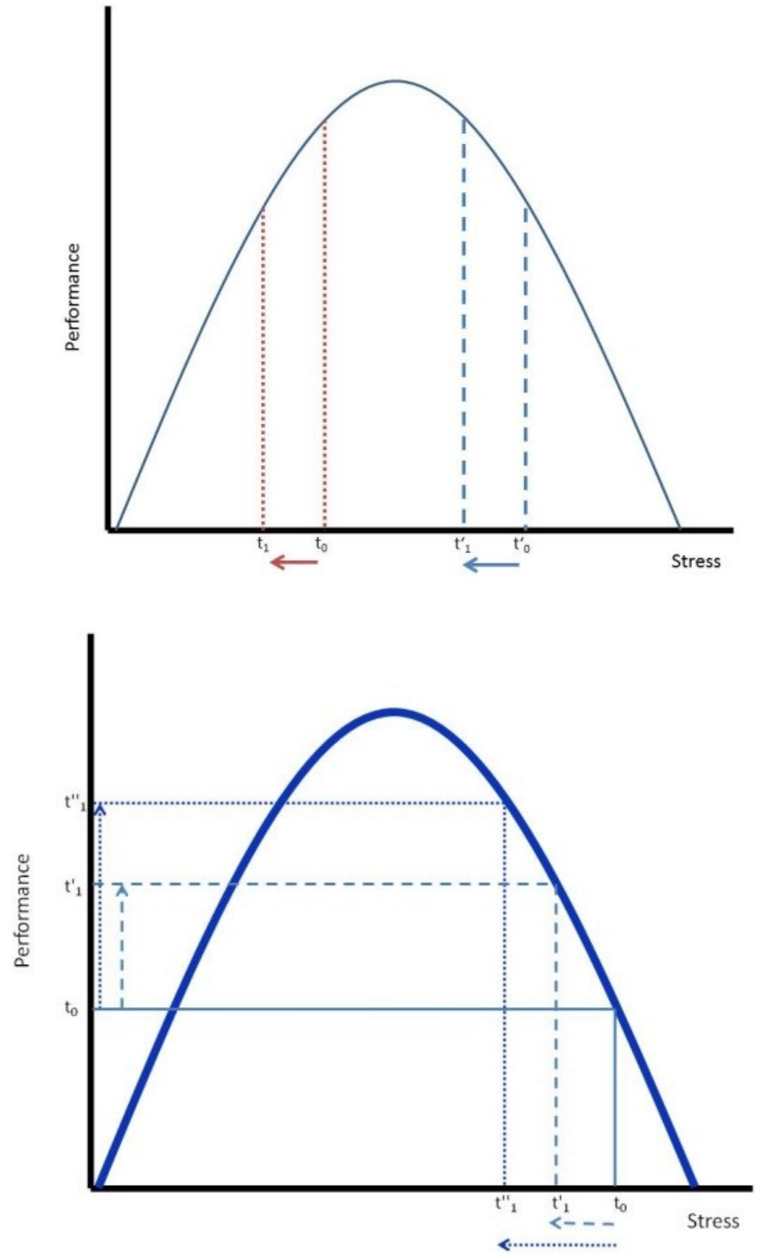
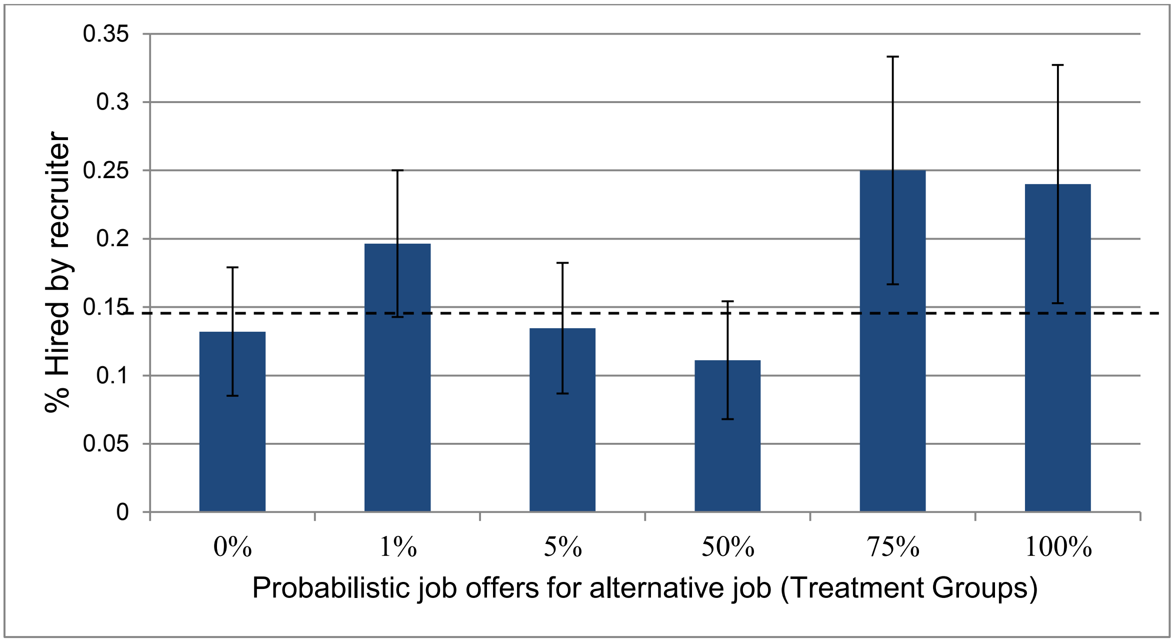


Figure 1.7: Fraction employed by recruiter by treatment group



Note: The dotted line represents the fraction that would have been hired in the absence of the experiment.

Figure 1.8: Average standardized test score by treatment group and mental health status

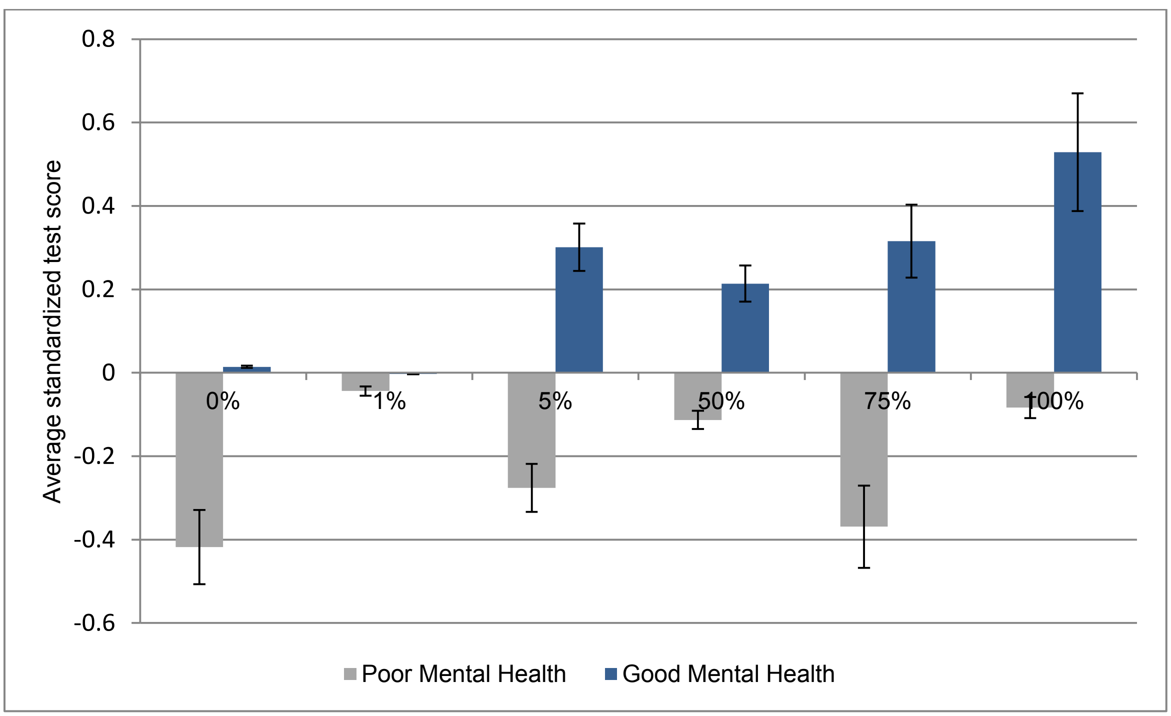


Figure 1.9: Fan regression of difference between guaranteed outside option and no outside option across the mental health distribution

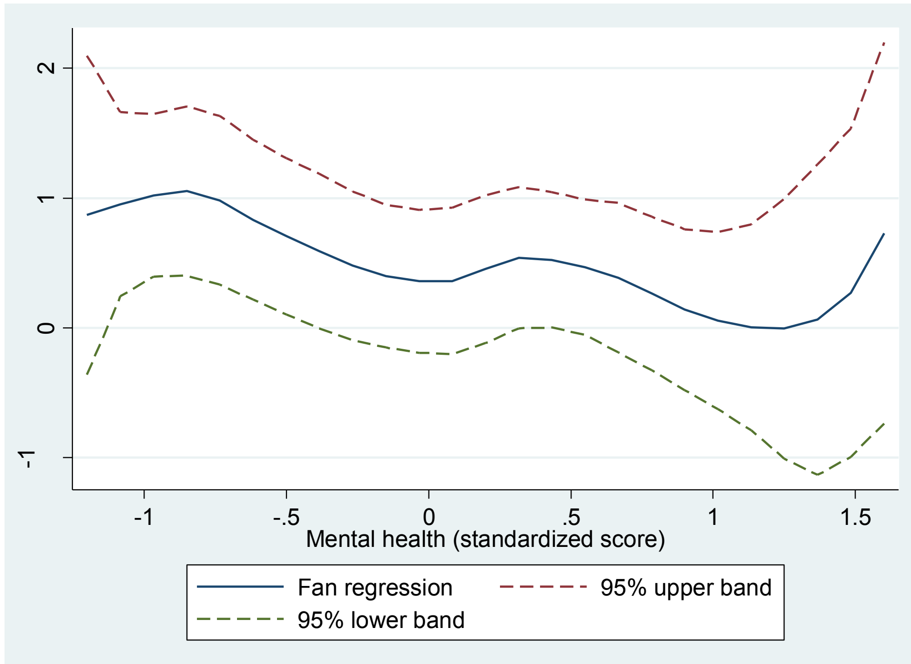


Figure 1.10: Average employment by recruiter by treatment group and mental health status

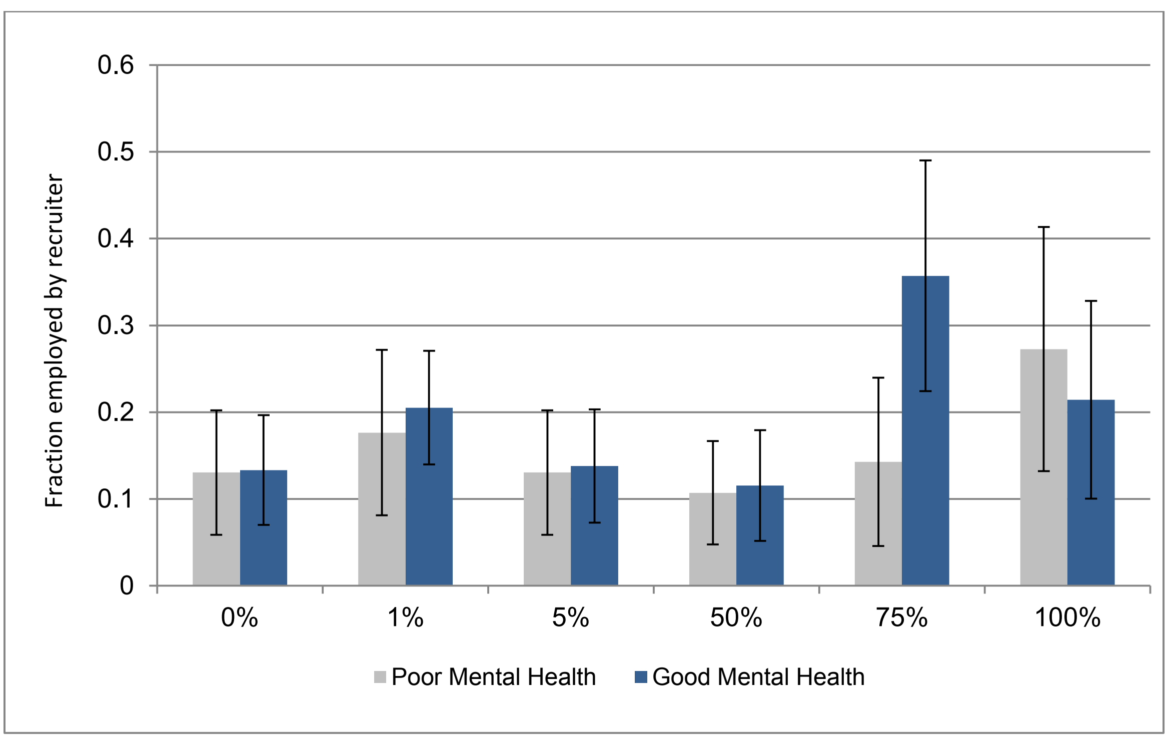


Figure 1.11: Average performance by treatment group and baseline ability

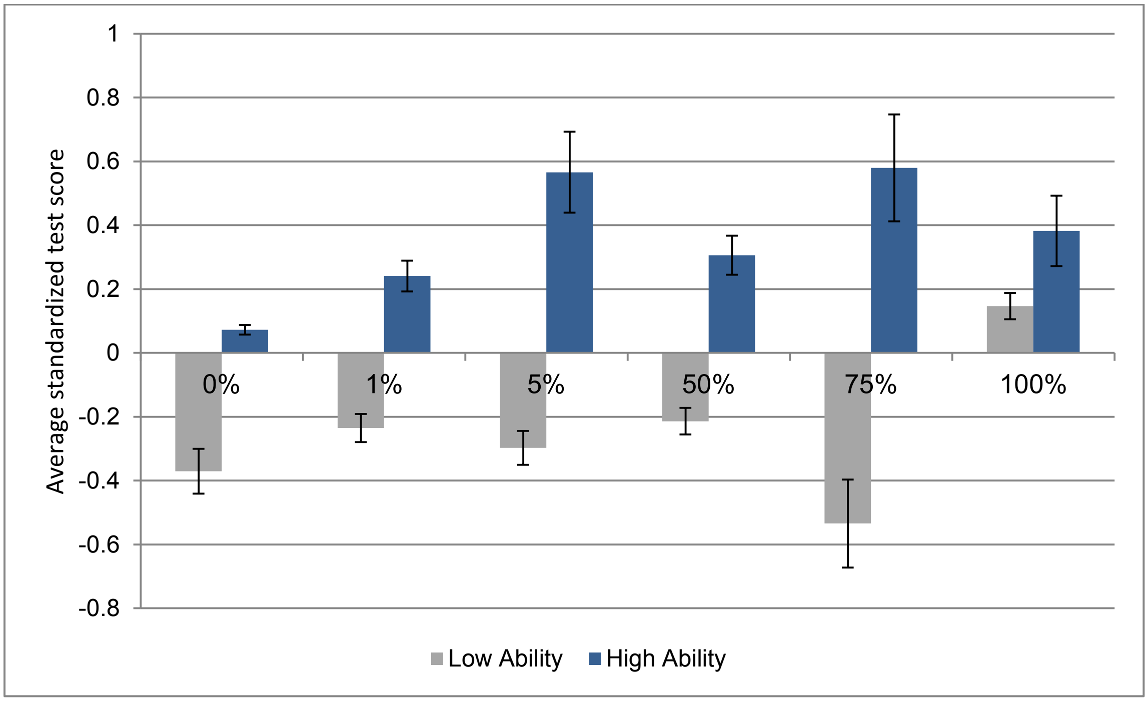


Figure 1.12: Fan regression of difference between guaranteed outside option and no outside option across baseline ability distribution

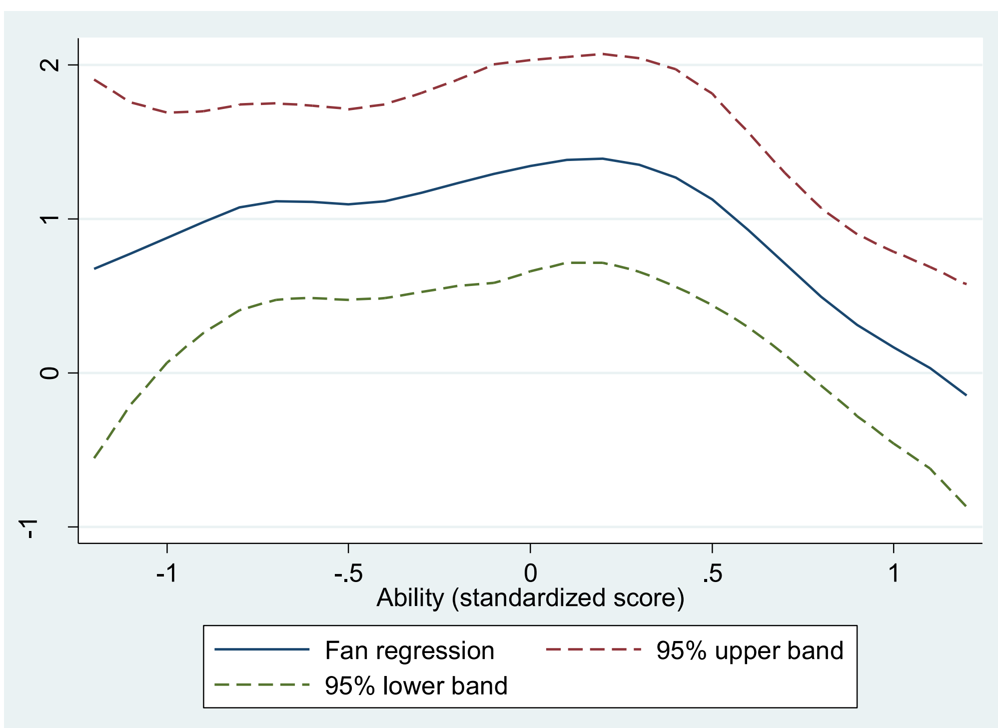


Figure 1.13: Average standardized test score by treatment group and baseline ability

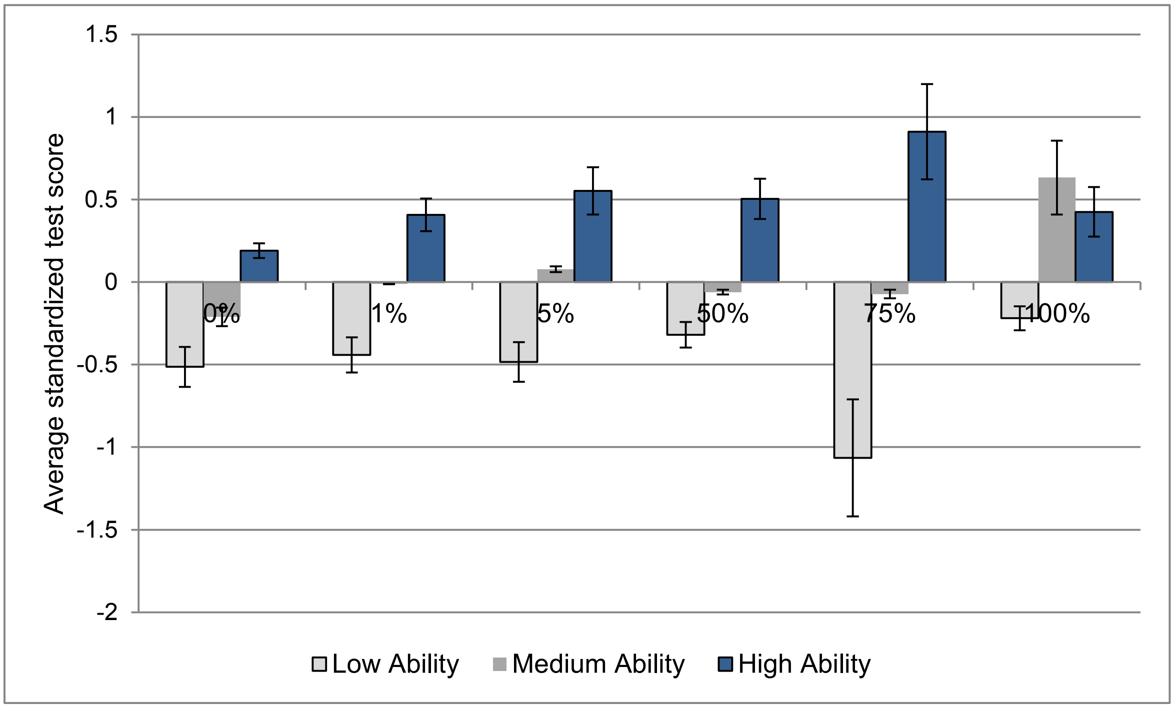


Figure 1.14: Average employment by recruiter by treatment group and baseline ability

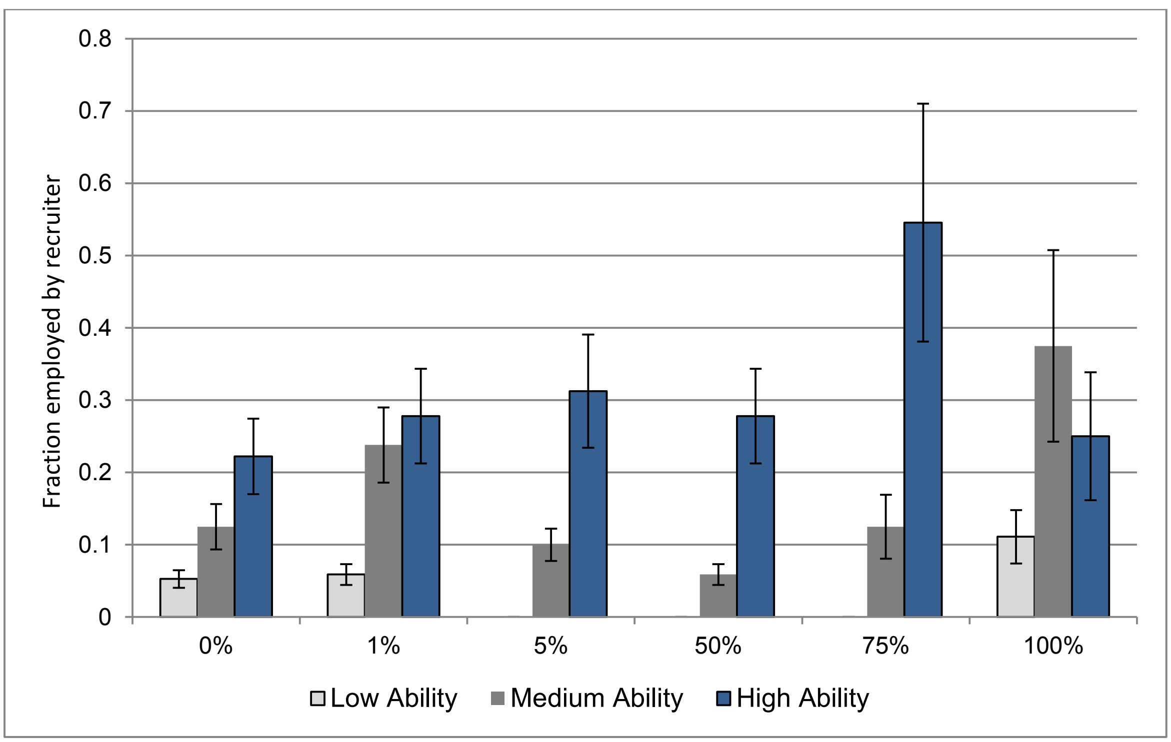


Table 1.1: Sample and attrition

Panel A: Sample (pre-treatment): Treatment Assignment

		All	0%	1%	5%	50%	75%	100%
Sample frame	N	278	55	56	56	56	28	27
(Intended)	%		0.198	0.201	0.201	0.201	0.101	0.097
Main Sample	N	268	53	56	52	54	28	25
(Actual)	%		0.198	0.209	0.194	0.201	0.104	0.093

Panel B: Training Participation and Survey Data Completion

	Administrative Data	Survey Questionnaires		
	Attended training	Pre-treatment	Post-treatment	
	Every day	Baseline	At least once	Every day
	(1)	(2)	(3)	(4)
0% Job Guarantee	0.906	0.981	0.906	0.830
	[0.041]	[0.019]	[0.041]	[0.052]
1% Job Guarantee	0.964	0.946	0.964	0.946
	[0.025]	[0.030]	[0.025]	[0.030]
5% Job Guarantee	0.923	0.942	0.981	0.865
	[0.037]	[0.033]	[0.019]	[0.048]
50% Job Guarantee	0.944	1.000	0.944	0.870
	[0.032]	[0.000]	[0.032]	[0.046]
75% Job Guarantee	0.964	1.000	0.964	0.893
	[0.035]	[0.000]	[0.035]	[0.059]
100% Job Guarantee	0.960	1.000	1.000	0.960
	[0.040]	[0.000]	[0.000]	[0.040]
Avw of dep variable	0.940	0.973	0.955	0.888
N	268	268	268	268
p-values of F-tests:				
All (jointly equal)	0.810	**0.068**	**0.031**	0.221
0% and 1%	0.220	0.334	0.220	**0.055**
0% and 100%	0.339	0.319	**0.021**	**0.049**
1% and 100%	0.927	**0.080**	0.156	0.786
50% and 100%	0.759	.	**0.079**	0.142
75% and 100%	0.936	.	0.315	0.346

Notes:

The sample frame consists of 278 participants that were short-listed for training by the recruiter.

Panel A shows intended and actual assignment of the job probabilities. These distribtuion differ due to 10 participants that opted out of the research study (prior to learning their treatment status) or opted out of the training prior to the commencement of training. The main sample used in this paper consists of 268

Panel B presents average participation rates in training and survey data completion rates by treatment group. A partial set of p-values from pair-wise comparisons of treatment group means are presented. All those that are not presented have p-values greater that 0.10. The full set of results is available on request.

Table 1.2: Summary statistics and balancing tests

Baseline Characteristics:	**N (1)**	**0% (2)**	**1% (3)**	**5% (4)**	**50% (5)**	**75% (6)**	**100% (7)**	**F-stat[1] (8)**
		Treatment Assignment						
Demographics:								
Age	268	25.887	25.893	24.865	25.463	26.464	25.240	0.757
		[-5.176]	[-4.735]	[-4.334]	[-3.490]	[-5.903]	[-4.612]	
Married	268	0.189	0.250	0.135	0.093	0.250	0.120	0.207
		[-0.395]	[-0.437]	[-0.345]	[-0.293]	[-0.441]	[-0.332]	
# of children	250	0.388	0.431	0.277	0.132	0.560	0.200	0.154
		[-0.909]	[-0.922]	[-0.743]	[-0.520]	[-1.083]	[-0.577]	
Income (in USD, 3 months)	225	181.72	247.12	167.54	199.88	294.96	282.83	0.240
		[-203.65]	[-272.13]	[-187.69]	[-203.72]	[-299.11]	[-342.76]	
Education, Ability and Experience:								
Years of schooling	268	13.264	13.071	13.115	13.130	13.107	13.600	0.277
		[-0.858]	[-0.931]	[-1.041]	[-0.953]	[-0.786]	[-1.000]	
Ability (standardized)	268	-0.075	-0.006	-0.020	0.034	0.116	0.010	0.978
		[-0.960]	[-1.021]	[-0.989]	[-1.063]	[-0.992]	[-1.013]	
Ever worked	268	0.906	0.857	0.750	0.944	0.929	0.840	0.083
		[-0.295]	[-0.353]	[-0.437]	[-0.231]	[-0.262]	[-0.374]	
Worked last month	252	0.600	0.647	0.638	0.577	0.536	0.792	0.357
		[-0.495]	[-0.483]	[-0.486]	[-0.499]	[-0.508]	[-0.415]	
Any work in past 6 months	252	0.780	0.902	0.894	0.808	0.893	0.958	0.137
		[-0.418]	[-0.300]	[-0.312]	[-0.398]	[-0.315]	[-0.204]	
Months worked (max. 6)	252	2.820	2.922	2.468	2.538	2.429	3.083	0.759
		[-2.371]	[-2.226]	[-2.155]	[-2.313]	[-2.116]	[-2.225]	
p-values associated with F-tests for joint significance of covariates[3]:								
Compared to all other groups			0.175	0.395	0.400	**0.060**	0.146	0.223
Compared to 0%				**0.006**	0.397	**0.098**	0.210	0.014
Compared to 1%					0.782	**0.009**	0.559	0.147
Compared to 5%						0.468	0.405	0.772
Compared to 50%							**0.078**	**0.025**
Compared to 75%								**0.004**

Notes:

The table reports group means or proportions (where applicable, e.g. married). Standard deviations are reported in parentheses. The main sample of 268 participants is used here. Data used here comes from both the baseline self-administered questionnaire and administrative data collected by the recruiter. Income is measured in USD and includes all self-reported income from the last three months including the following explicit categories: Farming; Ganyu (piece-work); Formal employment; Own business; Remittances; Pension; and Other. The ability scores are a composite measure of literacy and numeracy scores and are presented in standardized units. See Figures 1.3a, 1.3b and 1.3c for the distribution of these scores.

[1] These p-values correspond to the joint F-test of the means/proportions being equal across all treatment groups.

[2] This refers to the number of pairwise comparisons between treatment groups that are statistically significant at the 5 percent level. A total of 15 comparisons are made for each variable.

[3] These F-statistics report the p-value from the joint F-test for whether all the covariates listed are jointly equal in predicting assignment to the treatment group.

Table 1.3: Average performance on training tests by treatment group

	Average training test score		
	(1)	(2)	(3)
0% Job Guarantee	-0.176	-0.19	-0.177
	[0.147]	[0.142]	[0.142]
1% Job Guarantee	-0.015	-0.009	-0.005
	[0.136]	[0.126]	[0.126]
5% Job Guarantee	0.041	0.066	0.04
	[0.132]	[0.113]	[0.119]
50% Job Guarantee	0.041	0.039	0.031
	[0.124]	[0.119]	[0.122]
75% Job Guarantee	-0.039	-0.037	-0.028
	[0.241]	[0.209]	[0.207]
100% Job Guarantee	0.259	0.261	0.261
	[0.195]	[0.200]	[0.198]
Observations	258	258	258
R-squared	0.01	0.19	0.2
Stratification cell fixed effects?	No	Yes	Yes
Includes controls?	No	No	Yes
p-values of F-tests:			
0% and 100%	0.076	0.069	0.073

Notes:

This table presents mean performance on the recruiter adminstered training tests by treatment group. The average standardized test score is constructed by taking the average of the standardized test score from the three tests. Individual tests are standardized by using the sample mean and standard deviation for the relevant test. Treatment status was randomly allocated and stratified by ability quintile and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Table 1.4: Average performance (engagement in training) by treatment group

	Engagement in training (contributions)					Performance Index
Dependent Variable:	**Any**	**# total**	**# good**	**# neutral**	**# bad**	**Performance Index**
	(1)	(2)	(3)	(4)	(5)	(6)
0% Job Guarantee	0.649	1.503	0.528	0.612	0.363	-0.118
	[0.071]	[0.226]	[0.099]	[0.130]	[0.108]	[0.080]
1% Job Guarantee	0.608	1.574	0.795	0.585	0.195	-0.006
	[0.067]	[0.259]	[0.134]	[0.126]	[0.090]	[0.079]
5% Job Guarantee	0.723	1.604	0.690	0.705	0.209	0.043
	[0.063]	[0.220]	[0.156]	[0.129]	[0.059]	[0.081]
50% Job Guarantee	0.641	1.377	0.767	0.386	0.224	-0.060
	[0.069]	[0.212]	[0.135]	[0.095]	[0.064]	[0.069]
75% Job Guarantee	0.720	1.258	0.705	0.480	0.072	-0.004
	[0.087]	[0.232]	[0.155]	[0.123]	[0.050]	[0.091]
100% Job Guarantee	0.761	2.247	0.938	1.035	0.273	0.251
	[0.082]	[0.418]	[0.193]	[0.244]	[0.112]	[0.134]
Observations	262	268	268	268	268	268
R-squared	0.690	0.493	0.415	0.354	0.170	0.078
Stratification cell FE's?	Yes	Yes	Yes	Yes	Yes	Yes
Includes controls?	Yes	Yes	Yes	Yes	Yes	Yes
p-values of F-test:						
0% and 100%	0.310	0.119	0.058	0.127	0.571	0.018

Notes:

This table presents mean performance as measured by engagement recorded by the recruiter by treatment group. "Any contribution" is a binary indicator if the job trainee ever engaged verbally in training. The "total number of contributions" is the cumulative number of contributions made by the job trainee during the three days of training, and then separated out by quality as determined by the recruitment staff. The performance index is a summary measure of the performance indicators. It is constructed by taking the average of the normalized values of "Average test score", "Any contribution", "Total number of contributions", "Number of good contributions", "Number of neutral contributions", and "Number of bad contributions". Treatment status was randomly allocated and stratified by quintile ability and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are reported.

Table 1.5: Mean effort by treatment group

Dependent Variable	Ever late	Always late	Mins early or late	Studied (Hours)	Radio/TV (Hours)	Effort index
	Administrative Data			Survey Data		
	(1)	(2)	(3)	(4)	(5)	(6)
0% Job Guarantee	0.183	0.017	-24.400	1.179	1.155	0.214
	[0.053]	[0.020]	[2.156]	[0.131]	[0.123]	[0.083]
1% Job Guarantee	0.185	0.001	-21.405	1.148	1.582	0.000
	[0.052]	[0.003]	[1.856]	[0.110]	[0.132]	[0.079]
5% Job Guarantee	0.321	0.020	-19.187	0.951	1.356	-0.088
	[0.065]	[0.021]	[2.394]	[0.100]	[0.160]	[0.090]
50% Job Guarantee	0.175	0.019	-21.747	1.096	1.512	0.017
	[0.056]	[0.020]	[2.146]	[0.099]	[0.133]	[0.069]
75% Job Guarantee	0.254	0.039	-19.846	1.139	1.408	0.026
	[0.087]	[0.039]	[3.177]	[0.140]	[0.166]	[0.118]
100% Job Guarantee	0.276	0.080	-19.179	0.750	2.037	-0.373
	[0.091]	[0.055]	[4.153]	[0.079]	[0.247]	[0.144]
Observations	259	259	259	254	254	259
R-squared	0.270	0.070	0.657	0.689	0.707	0.104
Stratification cell FE's?	Yes	Yes	Yes	Yes	Yes	Yes
Includes controls?	Yes	Yes	Yes	Yes	Yes	Yes
p-values of F-test:						
0% and 100%	0.340	0.271	0.247	0.005	0.002	0.001

Notes:

This table presents the average effort by treatment group using both administrative data and survey data. "Always late" is a binary indicator equal to 1 if the job trainee always arrived late for training. "Ever late" is a binary indicator equal to 1 if the job trainee ever arrived late to training. "Minutes early/late" is a continuous variable recording the minutes early (-) or late (+) job trainees arrived at training. Time use in columns 4 and 5 comes from survey data and is the average hours reported by respondents across the 3 observations for each activity. The effort index is a summary measure of the effort indicators. It is constructed as the average of the normalized values of: "Minutes early/late", " Hours studying training materials", "Hours watching television/listening to the radio". Treatment status was randomly allocated and stratified by ability quintile and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Table 1.6: Alternative explanations?

Average	Food Expenditures (in MKW)	Eat out Expenditures (in MKW)	Perceived chance of employment with recruiter
	(1)	(2)	(3)
0% Job Guarantee	349.479	124.151	73.058
	[77.118]	[16.339]	[3.557]
1% Job Guarantee	425.084	165.495	73.538
	[98.487]	[15.067]	[2.996]
5% Job Guarantee	372.697	154.952	76.109
	[92.836]	[21.179]	[3.170]
50% Job Guarantee	439.111	147.49	72.706
	[97.689]	[20.097]	[2.343]
75% Job Guarantee	335.364	183.878	83.596
	[74.342]	[27.507]	[3.376]
100% Job Guarantee	328.482	123.887	77.596
	[79.742]	[23.159]	[3.553]
Observations	256	256	256
R-squared	0.36	0.6	0.94
p-values of F-test:			
0% and 100%	0.797	0.543	0.363

Notes:

This table presents the treatment group means for each outcome.

Food Expenditures (in MKW) is the average amount spent on food reported by the respondent across the 3 training days. "Eat out expenditures (in MKW)" is similar except measures food expenditures for food consumed away from the home.

"Perceived chance of employment with recruiter" is constructed using the following question: “What percentage chance do you think you have of getting one of the available positions for the RECRUITER’S PROJECT?” with the following options: No chance of getting a job; Less than 25 percent; Between 25 and 50 percent; 50 percent; Between 50 and 75 percent; Between 75 and 99 percent; and Certain about employment with recruiter. To create a measure of the likelihood of employment I assign the mid-point to categories that are brackets and creating a continuous variable.

Treatment status was randomly allocated and stratified by ability quintile and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Table 1.7: Employment (with recruiter) by treatment group

Panel A: OLS Regressions						
	Full sample			**Attend all training days**		
	(1)	(2)	(3)	(4)	(5)	(6)
0% Job Guarantee	0.132	0.126	0.133	0.137	0.129	0.136
	[0.047]	[0.045]	[0.046]	[0.049]	[0.047]	[0.047]
1% Job Guarantee	0.196	0.195	0.197	0.2	0.198	0.197
	[0.054]	[0.051]	[0.051]	[0.055]	[0.052]	[0.052]
5% Job Guarantee	0.135	0.139	0.136	0.137	0.143	0.141
	[0.048]	[0.043]	[0.044]	[0.049]	[0.043]	[0.045]
50% Job Guarantee	0.111	0.114	0.108	0.118	0.117	0.11
	[0.043]	[0.043]	[0.044]	[0.046]	[0.046]	[0.047]
75% Job Guarantee	0.250	0.241	0.238	0.259	0.256	0.251
	[0.083]	[0.074]	[0.071]	[0.085]	[0.075]	[0.073]
100% Job Guarantee	0.240	0.250	0.256	0.240	0.25	0.255
	[0.086]	[0.091]	[0.089]	[0.086]	[0.091]	[0.089]
effects?	No	Yes	Yes	No	Yes	Yes
Includes controls?	No	No	Yes	No	No	Yes
Observations	268	268	268	260	260	260
R-squared	0.18	0.28	0.29	0.18	0.29	0.3
p-values of F-tests:						
0% and 100%	0.274	0.224	0.221	0.314	0.241	0.238

Panel B: Probit Regressions

	Full sample			Attend all training days		
	(1)	(2)	(3)	(4)	(5)	(6)
0% Job Guarantee	-0.088	-0.095*	-0.090*	-0.085	-0.094*	-0.090*
	[0.065]	[0.054]	[0.051]	[0.068]	[0.056]	[0.054]
1% Job Guarantee	-0.034	-0.048	-0.051	-0.032	-0.048	-0.052
	[0.075]	[0.064]	[0.060]	[0.077]	[0.067]	[0.062]
5% Job Guarantee	-0.085	-0.093*	-0.093*	-0.085	-0.094*	-0.093*
	[0.065]	[0.052]	[0.049]	[0.068]	[0.055]	[0.052]
50% Job Guarantee	-0.106*	-0.104**	-0.109**	-0.103	-0.104*	-0.109**
	[0.061]	[0.051]	[0.046]	[0.064]	[0.053]	[0.049]
75% Job Guarantee	0.008	-0.024	-0.031	0.015	-0.014	-0.022
	[0.094]	[0.073]	[0.067]	[0.099]	[0.080]	[0.073]
effects?	No	Yes	Yes	No	Yes	Yes
Includes controls?	No	No	Yes	No	No	Yes
Observations	268	268	268	260	260	260

Notes:

Panel A presents employment rates (with recruiter) by treatment group.

Panel B presents the partial derivative at the mean of the covariates of employment of the 0-, 1-, 5-, 50-, 75- job probabilities treatment compared to the 100 percent treatment group where employment risk is 0.

Columns 1 through 3 present results for the full sample, while Columns 4 through 6 exclude those that did not attend all training days.

Treatment status was randomly allocated and stratified by quintile ability and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented. *** indicates significance at the 1% level, ** indicates significance at the 5% level, * indicates significance at the 10% level

Table 1.8: Heterogeneity in test performance

Dependent Variable: Standardized average test score	**Mental health (standardized)**	**Mental health (above)**	**Ability (standardized)**	**High Ability**
	(1)	(2)	(3)	(4)
0% Job Guarantee	-0.213	-0.418	-0.157	-0.371
	[0.153]	[0.237]	[0.141]	[0.199]
1% Guarantee	0.099	-0.044	-0.009	-0.235
	[0.153]	[0.225]	[0.130]	[0.194]
5% Guarantee	-0.036	-0.276	0.054	-0.298
	[0.159]	[0.222]	[0.117]	[0.162]
50% Job Guarantee	-0.017	-0.113	0.020	-0.214
	[0.143]	[0.152]	[0.113]	[0.152]
75% Job Guarantee	0.213	-0.369	-0.106	-0.534
	[0.291]	[0.312]	[0.189]	[0.301]
100% Job Guarantee	0.211	-0.084	0.257	0.147
	[0.205]	[0.294]	[0.191]	[0.347]
0% Job Guarantee X Het	0.273	0.510	0.327	0.443
	[0.111]	[0.297]	[0.155]	[0.291]
1% Guarantee X Het	0.193	0.269	0.314	0.476
	[0.121]	[0.306]	[0.100]	[0.266]
5% Guarantee X Het	0.379	0.679	0.449	0.864
	[0.150]	[0.305]	[0.106]	[0.235]
50% Job Guarantee X Het	0.105	0.243	0.338	0.520
	[0.155]	[0.315]	[0.123]	[0.241]
75% Job Guarantee X Het	0.680	1.030	0.836	1.114
	[0.329]	[0.602]	[0.167]	[0.435]
100% Job Guarantee X Het	0.368	0.621	0.264	0.235
	[0.251]	[0.409]	[0.156]	[0.381]
Observations	202	202	258	258
R-squared	0.120	0.092	0.195	0.117
p-values of F-tests:				
$\beta_1 = \beta_6$	0.133	0.187	0.176	0.243
$\beta_1 + \beta_7 = \beta_6 + \beta_{12}$	0.099	0.377	0.083	0.198

Notes:

This table presents treatment group means and their interaction with different baseline covariates. Treatment status was randomly allocated and stratified by ability quintile and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Chapter 2

Employment Exposure: Experimental Evidence on Employment and Wage Effects

2.1 Introduction

Understanding the key determinants of wage growth can inform policy interventions that reduce poverty. A number of competing wage determination theories exist. First, a long empirical literature supports the notion that educational attainment is an important determinant of wages. Second, both firm-specific and general work experience are particularly important determinants of wage growth. Lastly, the quality of a job-match is an important determinant of wage growth. There is debate, however, about whether job turnover increases or decreases wages. One hypothesis is that job-turnover results in reduced wages; an alternative claims that job-shopping may result in a better match and therefore higher wages. The empirical literature finds stronger support for the latter.[48]

While the determinants of wages have been extensively studied in the United States and other developed country settings, much less evidence exists for developing countries. One exception to this is the extensive literature examining the returns to schooling.

[48] The empirical literature on the effects of job stability/mobility is extensive. Devine and Kiefer (1991) provide a review of this literature.

This literature shows that the returns to schooling are larger for females, and in countries with lower GDP per capita (Psacharopoulus, 1973, and 1994). An explanation for the higher returns is the scarcity of skilled labor (Mwabu and Schultz, 2000). Heterogeneous initial conditions in terms of the stock of skilled labor and other factors that affect the productivity of labor may imply that the determinants of wages differ across countries. Identifying the importance of factors such as experience, tenure and job mobility (or stability) is therefore important to understanding wage growth in different contexts. In this paper, I study the effect of experience on employment and wages in urban Malawi.

A key challenge in estimating the impact of work experience is that experience is endogenous and likely to be correlated with other factors that affect employment or wages. For example, individuals who acquire work experience may exhibit better non-cognitive skills not observable in the data. Several papers have shown that non-cognitive influence labor market outcomes (Bowles, Gintis, and Osborne 2001; Jacob 2002; Heckman, Stixrud and Urzua, 2006). Because so many characteristics are likely correlated both with past acquired work experience and future labor market outcomes, the assumptions for selection on observables are unlikely to be satisfied even when high quality survey or administrative data are available. Estimating the returns to work experience in developing countries is further constrained by the dearth of detailed labor force and panel data, particularly in Africa. Even though the prevalence of labor force panel studies is increasing they often lack detailed retrospective employment histories or sufficient detail on jobs to accurately measure acquired work experience. To circumvent this data limitation, most existing studies use a measure of "potential experience" that is the difference between an individual's age and his years of schooling in estimating the employment and wage effects of experience. However, the prevalence of interrupted or delayed schooling and periods of unemployment renders potential experience a poor proxy for actual experience in developing countries.

In this paper I overcome the identification challenge by exploiting an unusual source of random variation in short term employment. I also collect data that contain more detailed information about employment history than typically available, and measure actual rather than potential work experience. The exogenous variation I exploit derives from an experimental study conducted in Malawi and discussed in detail in Godlonton (2013). Specifically, job-trainees were randomly allocated a probabilistic chance of short term employment in a real job. There were six treatment groups. Individuals were assigned to receive a 0-, 1-, 5-, 50-, 75- or 100-percent chance of employment in research assistance activities at the completion of the training and recruitment process (even if they were not hired by the recruiter). These probabilistic chances of jobs can be used as an instrument for short term work experience. I have rich baseline data, including a baseline survey and resume for each of the 268 job trainees. Outcome data come from a follow-up survey that collects data on retrospective work histories for the eight month period following the experiment.

By instrumenting for an individual's work experience using his randomly-assigned chance of gaining experience from the short term job, I am able to estimate the effect of short term work experience on employment and job search strategies. First, the estimated impact on employment after eight months is positive, though imprecisely estimated. Individuals offered an alternative job were between 10.6 and 13.9 percentage points more likely to be employed on average during the post-intervention. The estimated impact of experience on the probability of job search and the likelihood of holding multiple concurrent jobs across the eight month period following the intervention is positive but not statistically significant.

Second, I do find a sizeable wage return to work experience. Individuals who were assigned to receive work experience earn on average approximately $3.80 to $4.19 more per day, as estimated in specifications that do not condition upon employment. This is a large return representing a 75 to 83 percent increase in daily wages. In specifications that exclude the unemployed and use logged wages, the estimated effect is only somewhat smaller, with experience increasing wages by between 60 and 67 percent. Some of the increase in the wage may be attributed to an increase in the number of hours worked as this increases by approximately four hours per week (although the effect is not statistically different from zero). Another mechanism for the increase in wages is changes in occupation. I find that the short term research assistance experience prompts a shift away from agriculture and related occupations and towards clerical and related occupations. I examine a number of potential mechanisms through which experience causes wage increases. The data do not support the hypotheses that expanded social networks, signaling of ability from letters of reference, or increased reservation wages are behind the increase in wages. Indirect evidence is most consistent with the idea that experience facilitates skill acquisition, and skill is rewarded in the external labor market.

Furthermore, there is interesting heterogeneity in the employment and wage effects. Specifically, individuals of lower ability (as assessed by a numeracy and literacy test) benefit the most from the work experience. For this subgroup, the effect of experience on the probability of employment is statistically significant. Although the small sample size limits statistical precision, there is suggestive evidence that the employment effects are more are in fact growing over time for low ability types.

Overall, the results in this paper suggest substantial wage returns to even very limited work experience. The results are large when compared to non-experimental estimates that rely on variation in potential experience. However, making direct comparisons to the non-experimental estimates is difficult given the lack of variation in the amount of experience acquired for those induced to work by the experiment. The impacts are also large relative to experimental estimates of job training programs, which typically find modest effects at best (Heckman, Lalonde, Smith, 1999; Kluve, 2006). However, in this paper I study a very different context where the returns to experience may be significantly larger due to the scarcity of skills. Also, unlike most job training programs in developed countries, experience in this context is actually targeted to relatively skilled individuals, and individuals possibly gain general skills.

The paper is organized as follows. Section 2.2 provides background information both on related literature, relevant aspect of Malawian urban labor markets where this study is conducted and a description of the intervention. Section 2.3 describes the data used and Section 2.4 presents the empirical strategy. Section 2.5 presents and discusses the results. Section 2.6 concludes.

2.2 Context and data

2.2.1 Malawi: Education, experience and earnings in wage employment

Like much of Sub-Saharan Africa, the majority of Malawians depend primarily on subsistence agriculture. Internal migration to urban centers is high and rising (HDR, 2009), however. The trend towards urbanization means that understanding wage growth is particularly important in order to inform labor policies targeted to the growing urban labor force.

Previous studies of the return to education in Malawi estimate wage increases of between six and ten percent per additional year of schooling (Chirwa and Zgovu, 2001; and Chirwa and Matita, 2009). These estimates of returns to each additional year of schooling are consistent with relatively high point estimates of the effects of completing primary, secondary and tertiary (Psacaharopoulos and Patrinos 2002; Castel, Phiri and Stampini, 2001). One study has estimated the Mincerian return to experience for Malawi, finding that every additional year of potential experience is associated with a wage increase of approximately five percent (Chirwa and Matita, 2009). A five percent return to each year of experience is high relative to the marginal value of education in other countries; King, Montenengro and Orazem (2012) review Mincerian estimates of the return to experience from 122 datasets across 86 developing countries and find estimates between -1 and 4.25 percent per additional year of experience.

However, using potential experience as a measure of accumulated experience has been widely criticized, particularly in labor markets where there is high job turnover and general employment instability. Light and Ureta (1995) use work history data from the United States to show that specifications using cumulative experience and potential experience produce misleading estimates of the returns to tenure and experience in the United States. Using potential experience to measure work experience is particularly flawed in low-income countries due to high rates of grade repetition in school; exit and re-enrollment in schooling; and long spells of unemployment (Lockheed, Verspoor, et al. 1991; Lam, Ardington and Leibbrandt 2011; and Pugatch, 2013).

In this paper, I exploit the experimental variation from a randomized controlled trial conducted in urban Malawi discussed in greater detail in Godlonton (2013). The exogenous variation in work experience generated by that experiment provides the opportunity to examine the causal impact of a short term work opportunity on later labor outcomes.

2.2.2 Experimental variation

This paper makes use of the exogenous variation in work experience generated by a randomized controlled trial that offered individuals undergoing a real recruitment process a probabilistic chance of an alternative employment opportunity. Individuals were assigned a 0-, 1-, 5-, 50-, 75- or 100-percent chance of alternative employment in the event that they failed to secure employment through the recruiter's competitive hiring process. Thus, the probabilistic job guarantee provides a lower bound on the probability that an individual had the opportunity for employment at the conclusion of the recruiting process. The randomization was stratified by ability and prior experience with the recruiter. The alternative employment opportunity offered the same duration and wage as the standard employment offer from the recruiter. Individuals were still able to earn a job through the recruitment process by performing well during the job training, and those who secured both jobs were required to take the recruiter's job or turn down both job offers. Given that the recruiter's job and the alternative jobs were of equal duration and paid the same wage, those who became employed through the project acquired the same amount of work experience at the same pay whether they ultimately worked for the recruiter or in the alternative job. Estimation of the effect of the probabilistic job guarantee must account for the fact that the probabilistic jobs increased the likelihood of both being selected for the recruiter's job and being eligible for the alternative job (see Godlonton, 2013 for details).

The work experience acquired is short term. The job provided individuals with five days of paid work experience. The recruiter's job was for employment as an interviewer. The alternative jobs were different research assistant tasks, including archival research, data entry, and translation and transcription of qualitative interviews. Many of these tasks may embody some real acquisition of new and transferable skills for the participants. Upon completion of the job, participants a generic letter of reference.

Once the recruitment process was completed, the probabilistic chances of employment were realized. For individuals assigned a 1-, 5-, 50- or 75 percent chance of an alternative job; random draws were conducted. For example, an individual assigned a 75-percent chance of an alternative job drew a token from a bag that contained 75 red tokens and 25 green tokens. If the individual drew a red token then he was offered the alternative job; if he drew a green token, he was not. Similar draws were conducted by each individual, with token distributions adjusted for his randomly-assigned probabilistic treatment groups. Individuals assigned a 0-percent chance knew with certainty they were not eligible for alternative jobs and those assigned a 100-percent chance knew they were guaranteed alternative jobs, so no draws were conducted in those cases.

I use the treatment assignment (i.e. the probability of an alternative job) to instrument for acquired short term work experience. This unusual random determination of employment allows a unique opportunity to measure the causal effect of short term work experience on future labor market outcomes.

2.3 Data

Figure 2.1 outlines the timeline of the data used in this paper. The sample of respondents is drawn from a recruitment process hiring male interviewers, during which trainees also participated in an experiment that offered randomly determined probabilistic jobs. Data come from a baseline survey collected prior to the start of the recruitment process, administrative records about treatment assignment and employment realizations for both probabilistic alternative jobs and hiring by the recruiter, and a follow-up survey that was conducted nine months after the completion of the work opportunities presented by the experiment.

Baseline data: Prior to the start of the recruitment process, respondents completed numeracy and literacy tests and submitted their resumes. Using the numeracy and literacy scores I construct an ability measure. In addition to this information a baseline survey was conducted. The baseline survey collected information on basic demographics, general education and work experiences, as well as mental and physical health. The baseline survey was self-administered by respondents.

Probabilistic alternative job offers: I use both the assignment to treatment records, as well as the realization of the probabilistic draws (i.e. whether or not each participant was actually offered a job, conditional on the distribution he was randomly assigned to). Assignment to an employment probability was stratified by baseline ability quintile and prior experience with the recruiter. In Godlonton (2013) it is shown that the treatment assignment is balanced; in other words, there are no systematic differences in covariates between the different treatment groups.

Follow-up survey data: A follow-up survey was conducted nine months after the implementation of the experiment. While the reference period for the survey questions is the nine months following the completion of the work experience opportunities, some participants erroneously report work tied to the experiment. To deal with this survey recall error, I exclude the first month of recall data and rely only on the eight month period beginning one month after the completion of the work experience opportunities. The follow-up survey was conducted by phone and included an extensive module on job search, labor market perceptions (current and future likelihood of finding employment), current employment and employment experiences over the last eight months, current and past wages as well as a mental health module.

Table 2.1 shows that attrition was not statistically significantly associated with the treatment status. A total of 84.7 percent of the sample was successfully interviewed at follow-up. The attrition rate was lowest among participants who had received the 75-percent job guarantee (7.1 percent). Individuals assigned a 0-percent chance of an alternative job have the highest rate of attrition (18.9 percent). The difference in attrition between these two groups, although large, is not statistically significant (p=0.168). Moreover, the probability of receiving an alternative job does not predict the probability of being interviewed at follow-up (coeff. = 0.049, p-value = 0.433).

Table 2.2 shows that there is not differential attrition for other baseline characteristics including age, education, ability and previous work experience (Column 5). Respondents of the Ngoni tribe and those that had worked in the six months prior to baseline are slightly less likely to attrit (significant at the 5 percent level and 10 percent level respectively). However, these differences are not large in magnitude. Moreover, there is no systematic differential attrition by treatment status (i.e. the probability of the alternative job) that is correlated with baseline characteristics. To test this, I regress an indicator for being in the follow-up sample on the probability of being assigned an alternative job, the baseline characteristic of interest, and that probability interacted with the baseline characteristic (Appendix Table C.1).

The final analytical sample includes the 227 respondents found at follow-up. The average respondent in this sample is approximately 26 years old and 17.2 percent are married. Approximately 16.7 percent of the sample have at least one child, and of those that do have at least one child they have an average of 1.8 children. Respondents are relatively well educated for Malawi with an average of 13 years of education, but this is driven by the eligibility criteria of the recruiter which required individuals to have at minimum completed their secondary school education. Despite being relatively well-educated for Malawi all these men were actively seeking work at the time of the baseline sample and they reported earnings of only approximately $210 per month over three months prior to the experiment (Table 2.2, Column 2).

2.4 Empirical strategy

If experience was randomly assigned across individuals, then we could estimate the average treatment effect of experience on employment and wages using ordinary least squares (OLS). In that case, one would estimate the following regression equation:

$$y_i = \alpha + \beta_1 JO_i + X_i'\delta + \varepsilon_i \quad (1)$$

where y_i = employment (or wages) for individual i, T_i is a dummy indicator for whether or not the individual received a job, and X_i is a set of individual characteristics. However, in this setting work experience was not itself randomly assigned. Instead, individuals were randomly assigned different probabilities of obtaining work experience. These probabilistic job guarantees affected their likelihood of obtaining experience from one of two different types of jobs – the recruiter's job and the alternative job. I therefore implement an instrumental variables approach. The system of equations then estimated is:

$$Y_i = \alpha_0 + \beta_1 AnyJO_i + X_i'\delta + \varepsilon_i \qquad (3)$$

$$AnyJO_i = \pi_0 + \pi_1 P1_i + \pi_1 P5_i + \pi_1 P50_i + \pi_1 P75_i + \pi_1 P100_i + X_i'\varphi + \varepsilon_i \qquad (4)$$

where JO_i measures whether individual i was offered a short term job; $P1_i$, $P5_i$, $P50_i$, $P75_i$, $P100_i$ indicates the binary indicators for the treatment arms; and X_i represents a set of covariates. The set of covariates used is the same as those used in equation (2) and listed above. I also include stratification cell fixed effects to account for the fact that treatment assignment was stratified by ability and prior work experience with the recruiter. The key coefficient of interest is β_1. Y_i measures labor market outcomes of interest to examine both the intensive and extensive margins. To examine changes at the extensive margin I measure the impact the probability of being employed nine months after the experiment, and the fraction of months in which individuals are employed in the eight months following the intervention. To measure impacts at the intensive margin, I examine the average daily wage earned by individual i across that the eight month period. I allow for possible heteroskedasticity in the error terms by using heteroskedastic-robust standard errors.

For the probability of assignment to the alternative job to serve as a valid instrument for work experience, it needs to satisfy two conditions: i) the instrument must be correlated with the endogenous variable; ii) the probabilistic job offers must not affect later labor market outcomes except through the acquired work experience.

The first condition implies that assigned probability of alternative employment should predict whether or not the job-seeker acquired any job (recruiter or alternative job) through this intervention. Estimating the first stage relationship shows that the instrument is, indeed, relevant:

$$AnyJO_i = \pi_0 + \pi_1 P1_i + \pi_1 P5_i + \pi_1 P50_i + \pi_1 P75_i + \pi_1 P100_i + X_i' \varphi + \varepsilon_i \qquad (2)$$

In the equation above, $AnyJO_i$ is defined as a binary indicator equal to one if the respondent either received a randomly determined job or a recruiter's job. I use indicator variables for each of the treatment arms. $P1_i$ equals one if the individual received a one-percent probabilistic chance of a job, and $P5_i, P50_i, P75_i,$ and $P100_i$ are similar indicator variables for the 5-, 50-, 75- and 100-percent treatment arms. The omitted category is the group who received no chance of an outside job. X_i represents a set of covariates and includes: age, marital status, education dummies, a dummy indicator for whether the respondent has any children, the number of children that the respondent has, ability score (a composite measure of numeracy and literacy scores), dummy indicators for tribe, a dummy indicator if the respondent has any work experience, reports any work in the past month and any job search in the past month, and the number of months in the last six months he has worked.

Table 2.3 presents the first stage estimates. The first stage results show that the probabilistic jobs strongly predict the probability participants received any job (recruiter or alternative). This expected result derives mechanically from the assignment of alternative jobs, as well as through a behavioral response by participants to the job guarantees. As shown in Godlonton (2013) the probability of being hired by the recruiter was higher among those who received the 75- or 100- percent chance of an alternative job, likely because the improved outside option lowered stress and increased performance during the recruiting process. Both mechanisms work in favor of a higher probabilistic job guarantee causing a higher chance of subsequent employment. Table 2.3 Column 1 confirms this hypothesis. A total of 16.3 percent of individuals assigned a zero chance of an alternative job got a job. Individuals assigned a 1- or 5- percent chance of an alternative job are not more likely than those who were assigned a 0-percent chance to get any job. The coefficients are positive as predicted, though the standard errors are large. Individuals assigned a 50-, 75- and 100- percent chance of an alternative job are respectively 40.2, 56.8 and 83.7 percentage points more likely to get any job than those with no chance of the alternative job. The first stage F-statistic is 101.11, far above the rule of thumb threshold for weak instrument concerns. These results are robust to the inclusion of stratification cell fixed effects (column 2) and additional covariates (column 3).

The exogeneity condition for the IV strategy requires that, conditional on baseline characteristics, the probabilistic job offers do not affect later employment outcomes independently of acquiring a job through the experiment (recruiter or alternative). Monotonicity would have been violated if higher probabilistic job offers had reduced the likelihood of acquiring the recruiter's job. However, as shown in Godlonton (2013) this is not the case. In fact, individuals assigned a 75- or 100 –percent chance of an alternative job were about twice as likely to be hired by the recruiter as those who were not eligible at all for alternative jobs. A second concern is that the probabilistic job offers may have affected individuals' perceptions about their own ability to find employment. Results in Godlonton (2013) show that there is no effect of the probabilistic job offers on perception of ones' own likelihood of employment.

A third concern is that the probabilistic job offers affected skill acquisition during training, and that skill was subsequently rewarded by the labor market. The finding in Godlonton (2013) that individuals perform differentially on recruiter administered training tests during the recruitment process may initially heighten that concern. However, it is unlikely that there were general benefits to this training. The training conducted by the recruiter and evaluated in the performance tests was tailored to the specific needs of that particular recruiter's temporary job, interviewer positions for a health survey. Participants worked systematically through the questionnaire the recruiter planned to administer, in order to understand the terminology of and instructions for filling in each item. Participants were given systematic explanations about how to interpret questions, but the training was very specific to the survey in question. Skills related to this particular questionnaire are highly firm-specific and are unlikely to be marketable to the labor market. Moreover, for the training to have an impact in the labor market the differential performance of the participants needs to be observable to future employers. Individuals did not receive their grades on these assessment tests and letters of reference only described the nature of the job but not the employee's specific performance. As such, the only way for the differential performance during training to affect subsequent employment and earnings in the outside labor market after the intervention is for outside employers to value the specific content of the training conducted by the recruiter during the experiment. Given the nature of the recruiter's training, this is unlikely.[49] Generally, in this context when individuals apply for a new interviewer position even within the same firm they still are required to undergo the same training for each new survey as the content of each training and skills taught are specific to that survey. In other words, experienced and novice interviewers undergo the same training for each survey they work on.

Conditional on instrument validity, β_1 captures the local average treatment effect (LATE) of the short term job on labor market outcomes – employment and wages.

2.5 Results

Work experience may affect employment at the extensive margin, by changing the probability of employment, and the intensive margin, changing wages conditional on employment. In this section, I use the variation generated by the experiment to study the return to experience at each of these margins.

[49] I restrict the analysis by excluding those assigned the 100-percent treatment group; and those assigned the 0-percent treatment group. These sub-groups show that the results are slightly smaller and in some cases lose statistical significance which is not surprising as the sample sizes are small. These estimates also show that the results are not eliminated by dropping either of these groups which suggests that the results are not driven by differential learning (results not shown).

2.5.1. Returns to experience

Table 2.4 presents the impact of the short term work experience on job search, employment, and the concurrent number of jobs held. This table uses data aggregated by individual across the eight month post-intervention time period. The employment variable used is the probability of employment during this timeframe. This is constructed by calculating the fraction of months that the individual is employed over the eight months following the intervention. Similarly, the job search variable is defined as the average probability an individual actively sought work (whether or not they were employed). Again, like the employment variable this is constructed as the fraction of months an individual actively sought work in the post-intervention period. The measure of concurrent number of jobs held is constructed as the average number of concurrent jobs held during the last eight months.

Work experience increases the probability of employment by all three measures. The short term work experience provided by the experiment increased the probability of subsequent by 10.6 to 13.9 percentage points. The estimated coefficients increase in magnitude and precision when we include stratification cell fixed effects (column 2) and covariates (column 3). The estimated effect is large, representing a 25 to 33 percent increase in the probability of being employed. To explore the time dynamics behind the average effect estimated in Table 2.4, Figure 2.2 plots the estimated employment impacts of the job separately for each of the eight months following the intervention. Although the one-month estimates are imprecise, the effects are positive in each of the eight months and statistically different from one another.

Work experience also increases the probability of searching for a job (column 4) and the number of concurrent jobs held (column 7). These estimates are robust to including controls for stratification cell fixed effects and covariates.

Another margin along which employment may adjust is the number of days worked. Underemployment in Malawi is high, and there is plenty of scope to increase labor supply along the intensive margin. Data from a nationally representative household survey shows that urban men who have completed secondary school, the relevant comparison group for the experimental sample, work only 23.4 hours per week conditional on being employed. The follow up survey uses the standard labor supply survey instrument (2010/2011 IHS), so it measures hours of work rather than days of work in the past week. While I cannot measure the change in days of work, I can examine the change in the number of hours worked, and compute the implied average wage per hour. These results are also presented in Table 2.5. I find that among the employed, individuals are working approximately 40 percent more hours per week. In the local context, however, individuals are more likely to be able to adjust their labor supply at the daily than hourly margin, and they are paid per day rather than per hour. It is probably more accurate to interpret differences in hours as indicative of differences in the responsibilities of the job. Therefore, the results for hourly wage should be interpreted with caution. These estimates and show no statistically significant impact on the hourly wage (Table 2.5 columns 4 through 6). The magnitude of the coefficient indicates an increase of $0.72 per hour which is large in magnitude but it is not statistically significant.

Before turning to the mechanisms behind the increase in employment, Table 2.5 explores the impact of work experience on wages. The outcome measure is the individual's average daily wage over the eight-month follow up period. This measure is not conditional on employment, so periods when the individual is unemployed are included (as zeros) in the average. Daily wages – rather than the hourly wages used in much of the related literature – are the relevant unit in this context. Institutionally, all Malawian labor policies pertain to daily employment; for example, the minimum wage law is with respect to daily wages, not hourly wages. Daily or even more highly aggregated wages are also salient to respondents. The follow-up survey allowed individuals to choose the time unit for reporting their wages, with, 75.8 percent of respondents reporting monthly wages and 18.5 percent reporting daily wages. Therefore, while the literature about employment in developed countries uses hourly wages as the primary outcome of interest, daily wages are a more appropriate measure in this context.

Table 2.5 shows that individuals who gained work experience as a result of the experiment earn \$3.80 to \$4.19 more per day (Columns 7 through 9). This estimated effect is large relative to the average daily wage of approximately \$5.08 among the control group. The estimated impact represents a 75 to 83 percent increase in daily wages. As we did with the extensive margin effects, we can also consider the effect on wages separately for each of the eight months in the follow up period. Month-by-month estimates are plotted in Figure 2.3. In all months, the effect on daily wages is positive; it ranges between approximately one and six dollars.

The estimated wage impacts are surprisingly large and deserve further discussion. First, these results are not conditional on being employed; the outcome measure incorporates periods of unemployment as wages of zero. Therefore, part of the increase in wages is attributable to the gains in employment as shown in Table 2.4. Logged wages drops the unemployed, these results are present in Table 2.5 columns 10 through 12. The positive wage results persist, but are as expected the estimated coefficients are smaller in magnitude. However it is still large - the impact on the daily wage is 60 to 67 percent. These large point estimates are not driven by outliers. Figure 2.4 documents the wage distributions for those who did and did not receive a job and shows that the wage distribution among those who received a job is shifted to the right.

2.5.2 Mechanisms

Understanding the mechanisms may be helpful in reconciling the effects in this experiment with the much smaller effects estimated from non-experimental Mincerian estimates in Malawi and other settings. I find that only five days of work experience results in a 57 to 63 percent increase in subsequent earnings. This is equivalent to approximately ten years of experience in the Malawi non-experimental estimates (Chirwa and Matita, 2009). There are many reasons why the non-experimental estimates may be substantially smaller. First, the non-experimental study also uses an inferior measure of work experience. Potential experience overstates the amount of accumulated experience (considerably) in this context. Second, the type of experience studied by the experiment may be of higher quality than experience otherwise available to even educated Malawian men. While the experience provided through the experiment was short term, it was with a private, international employer. It is unlikely that five daysworth of work in the civil service will yield impacts similar to that observed here. Finally, the non-experimental estimates represent average returns to experience for a population that is less educated than the highly-skilled men included in the experiment. While the experimental subjects still experience frequent periods of unemployment, they may experience substantively different returns than a less educated counterpart.

There are many theoretical reasons to expect that experience (even short term informal work experience) leads to increased employment and wages. In this section, I discuss a number of these possibilities and discuss which might be most relevant in the current context. The particular mechanisms that I consider include changes in job search strategies or occupational choice; changes in contract type, altered social networks; skills acquisition; altered wage expectations; and human capital accumulation. The experimental setting was not designed to test these mechanisms directly. However, I present suggestive evidence against the backdrop of these outlined mechanisms, before turning an exploration of heterogeneity in the return to experience.

Shifts in occupation

One possibility is that individuals change their occupation if they are induced to receive a job. Using the retrospective calendar job histories, I classify each job according to the standard two-digit ILO occupation classification codes (using the ISCO-08 classification system). I then analyze employment in each industry separately, using three measures of occupation-specific employment. The first is a binary indicator for whether each individual ever worked in a given occupation. The second is the total number of months the respondent worked in each occupation. The third indicator is a binary for the respondent's modal occupation over the eight month follow up period.

In Table 2.6, each row reports the effect of work experience on employment in a separate occupation from an IV regression. The left hand panel corresponds to the binary ever-worked outcome; the middle panel is the number of months in the occupation; and the right hand panel is an indicator for modal occupation, as described above. Increased work experience as a result of the experimental variation caused increases in employment in the following occupations: administrative and managerial; and clerical and related worked. The same pattern is observed for the modal occupation held. Individuals were also more likely to have recent experience as professional, technical or related occupations but this pattern does not hold for the modal occupation. For clerical and related occupations the effect is large large, with the 13.1 percentage point increase in the probability of working in clerical or related occupations representing a 62 percent increase in the probability of employment in that field. Individuals appear to be switching from agriculture related, service and production and related occupations, but stronger claims are limited by the lack of statistical precision.

Employment contract type

Another mechanism through which experience may have affected wages is by altering the type of wage contract individuals secured after the intervention. Jobs vary in their duration, and short term positions are common in Malawi. I do not directly observe the duration of contracts in the follow up survey, but I can use information from the unit in which individuals reported their current job as a proxy for contract duration. Individuals self-reported the unit of payment for their current (primary) job at the daily, weekly, fortnightly or monthly level. I infer that lower-frequency reporting levels correspond to longer duration contracts, and construct a frequency of payment variable equal to one if the individual reports daily remuneration, two if weekly, three if fortnightly and four if monthly remuneration is reported. Table 2.7 reports effects of work experience on this proxy for job permanence. The estimated impact of work experience on payment frequency is -0.7. Individuals induced to receive work experience through the experiment appear to be working in less permanent positions. In this context, the change is consistent with higher wages, because wages for short term positions as research assistants or consultants on projects for international NGOs or donor agencies are often much higher than wages paid for the longer-term work offered by local employers or government agencies.

Social networks

Social networks have been touted as an important mechanism through which individuals acquire employment opportunities.[50] There are several theoretical reasons for why social connections are important in accessing employment. For the job-seeker, social connections can reduce search costs and lead to better quality matches (Calvo-Armengol, 2004; Mortensen and Vishwanath, 1994; Galeotti and Merlino, 2009).

Simply participating in jobs provided by this experiment may have facilitated new social connections between participants. These social connections may increase employment opportunities independently of the experience accrued. Unlike the experiments undertaken by Beaman and Magruder (2012) and Beaman et al. (2013) that are specifically set up to test various aspects regarding the role of social connections in job referrals, this experiment was not designed to induce variation in social connections or to test specific manner in which social connections might matter. However, I do measure the prevalence of social interactions that may have facilitated employment, such as whether individuals heard about job opportunities through individuals they met during

[50] See for example Beaman (2010) and Granovetter (1973).

the job opportunity, and whether the jobs they held during the eight month period following this job opportunity were a direct result of a referral.

Table 2.8 panel A shows that individuals who received work experience as a result of the experiment are 23.4 percentage points more likely to have heard about a work opportunity through someone they met during this intervention. However, while individuals claim to hear more about job opportunities, they are not more likely to secure employment through one of the new connections. Individuals are 12.6 percent less likely to report securing a job through someone they met during this intervention, but the estimate is not statistically significant at conventional levels.

In sum, while the broadened network does suggest a modest impact on information about job opportunities, this information does not translate into employment and therefore does not explain the effect of experience in this experiment.

Signaling

Another mechanism is signaling of worker quality to employers (Spence, 1973). In this case it is possible that employers do not infer any inherent value of the work experience on worker productivity, but merely interpret it as a signal of ability. Upon completion of the work experience all participants received a standard letter of reference, which described the job in general terms but did not provide information about individual-specific performance. Given that these letters came from an international employer, however, employers may value the letter as a signal of underlying ability, rather than certification of skills acquired through experience.

Table 2.8 panel B shows that those who received work experience as a result of the experimental treatment were actually 7.4 percentage points *less* likely to use the reference letter than to those who did not receive a job.[51] Therefore, employers would not have received any signal about worker ability from the reference letters, and these letters are unlikely to have contributed later labor market outcomes. However, it may still be possible that individuals put the work experience on their resume and this acts as a signal of ability.

Wage expectations

The job may have altered individuals' wage expectations and reservation wages, with implications for job search strategies, duration of unemployment, and match quality. The wages paid during this experiment may have been higher than reservation wages at baseline. If individuals updated their expectations by increasing their reservation wage, then the estimated impact on the employment effect might be muted, as individuals may be searching longer and differently for better paying jobs.

I examine this mechanism by looking at self-reported reservation wages. Table 2.8 Panel C presents the results from this exercise. The impact of receiving a job on the monthly reservation wage is $121.25, but it not statistically significant at conventional levels. More generally, the reported reservation wages are high, approximately 1.5 times higher than the average monthly income earned at baseline. Self-reported reservation wages also high relative to wages reported in the follow up survey. Transforming reported wages into full-time equivalent salaries with the assumption that individuals

[51] Individuals who received work in the alternative job and those who worked for the recruiter received reference letters as such it is possible that individuals who did not receive the randomly determined job used a reference letter. However, the large difference is not too surprising as a low fraction of those who received no alternative job offer worked for the recruiter, and therefore did not receive any reference letter that could be used for this purpose.

worked 20 days per month, then the average monthly wage earned at follow up was approximately $240, higher than at baseline but considerably lower than the reported reservation wage. While measurement error in the reservation wage complicates the interpretation of these results, there is no evidence that an increase in reservation wages is an important mechanism.

Human capital accumulation

A final potential mechanism is that individuals acquired skills attributable to the work experience induced by the experiment. Individuals who secured a job either worked as an interviewer or were assigned to data entry; data transcription or translation; or archival research jobs.

The results discussed in section 2.5.2 and presented in Table 2.6 show a change in occupational type. Individuals who received work experience are less likely to be employed in agriculture and more likely to be employed in clerical activities. Furthermore, individuals are 18.1 percentage points more likely to report having worked as a research assistant, the specific occupation in which they acquired experience. This is suggestive evidence that the work experience provided through the experiment generated occupation-specific skills that were rewarded by future employers.

While the data do not permit a direct test of the mechanism through experience increases which wages and employment, the indirect evidence suggests individuals may have acquired skills that are rewarded by the external labor market.

Heterogeneity

Understanding heterogeneous returns to work experience can help us interpret the large average effects and design policies to use work experience to improve employment outcomes. I explore heterogeneous returns by ability, work experience and education. To do so, I interact an indicator variable for having received an alternative job (JO_i) with the baseline characteristic of interest ($Base_i$*JO_i), using the set of treatment dummies as instruments for work experience. In this specification I instrument the endogenous regressors with the probability of an alternative job and this probability interacted with the baseline characteristic. Therefore, to examine the heterogeneity of the impacts I estimate the following set of equations:

$$Y_{it} = \alpha_0 + \beta_1 JO_i + \beta_2 Base_i + \beta_3 (JO_i * Base_i) + X_i'\delta + \varepsilon_{it} \qquad (5)$$

$$JO_i = \pi_0 + \pi_1 P_i + X_i'\varphi + \varepsilon_i \qquad (6)$$

$$(JO_i * Base_i) = \pi_2 + \pi_3 (P_i * Base_i) + X_i'\gamma + \varepsilon_i \qquad (7)$$

where: $Base_i$ is, in turn, the baseline ability score as determined by numeracy and literacy tests; a binary indicator for having completed college; and measures of current and cumulative labor market work experience.

Table 2.9 Panel A examines the heterogeneity of impacts by ability. To measure ability, I use test scores from a numeracy and literacy test administered to the respondents at baseline. I use a composite measure of ability that combines the numeracy and literacy test scores.[52] The estimated impacts are larger for individuals at the lower end of the ability distribution. To see this, consider an individual at the 25th percentile and the 75th percentile of the ability distribution. Individuals at the 25th percentile were 25 percentage points more likely to be employed if they were induced to receive job experience through the experiment, and they earn approximately \$11.01 more per day. On the other hand,

[52] The results are similar when using the numeracy and literacy scores separately.

individuals at the 75th percentile were 1.5 percentage points less likely to be employed, though they earn approximately $2.20 more per day.

Figure 2.5 plots the average post-treatment employment rate by month for low and high ability types. Individuals are classified as low ability if they scored below the mean on the composite literacy and numeracy test; and as high ability otherwise. The small sample limits the precision of the estimates by ability level, but the pattern is informative. The estimated impact on employment for low ability types is increasing over time, while there is no consistent pattern for the high ability types. The pattern for wages is relatively constant across the time period (not shown). This pattern of results suggests that the low ability types not only gain the most from the job but also that the employment returns are increasing over time.

Education and experience can serve as substitutes or complements in a Mincerian model. To examine the relationship in this context I consider heterogeneity by whether or not the respondent has a degree (Table 2.9 panel B). Due to sample restrictions imposed by the recruiter, the sample is composed entirely of individuals who have completed secondary schooling. Therefore, there is limited variation in educational attainment. The results show that the estimated impacts are largest for those without a university degree and are actually negative for those who have completed university.

Lastly, one possible reason that the estimated impacts are so large is that the experience provided in the experiment is the first job held by respondents. Table 2.9 Panels C and D explore the heterogeneity of the impacts with respect to work experience. Panel C uses recent job market attachment defined as whether the respondent was working a month prior to baseline; and Panel D uses an indicator for whether the individual has ever worked. Roughly 15 percent of the sample had no previous work experience. Perhaps surprisingly, the effects of work experience on subsequent employment do not differ by pre-experimental work experience.

2.6 Conclusion

This paper uses a novel experiment that generated exogenous variation in short term work experience in order to estimate the effect of such experience on employment in wages. The return to experience is large, with a 10.6 to 13.9 percentage point increase in post-intervention employment for those who received experience through the experiment relative to those who did not. Not only does experience increase the probability of being employed, but also, it has a sizeable effect on wages. Individuals who received work experience earn approximately 60 to 67 percent more per day than those who did not, with results concentrated among lower-ability individuals. This return to work experience is present in each of the eight months of the follow up period, and the average effects are larger than in previous estimates of the returns to experience in Malawi and other settings. Individuals shifted away from agricultural based occupations and into clerical and related work; they worked more hours per day and on contracts with shorter durations.

These results add to the policy debate about active labor market programs, which are designed to improve employment outcomes by providing participants with work experience. Proponents of work based programs believe that any job is a good job, and that getting a job will lead to job advancement and wage growth (Holcomb et al., 1998). However, the empirical evidence provides mixed results. In systematic reviews of the literature, the key take away is that the impact of job-training programs are modest at best (Heckman, Lalonde, Smith, 1999; Kluve, 2006). However, just like the returns to education, the impacts of such programs might be larger in low income countries. Betcherman, Olivas and Dar (2004) review the literature about impact evaluations of job training programs and find only 19 studies (none of which are in Africa) conducted in developing countries. In both this review and in another, by Nopo and Saavedra (2003) of the non-experimental literature in Latin America, the estimated impacts of job training programs appear to be larger in developing than developed countries.

The results may not be generalizable to a less skilled population within Malawi, or to a country whose underlying skill distribution and labor market conditions are different from Malawi. Even within Malawi, the treatment provided in the experiment is not available through any current public or private sector job training initiatives. Because the job opportunity provided within the experiment was of uniform duration, we cannot extrapolate from these results to the return to a longer period of experience. Lastly, the general equilibrium effects of such a program are not estimated. Given the small size of this intervention, it is not possible to determine if and the extent such a program if rolled-out would have on those individuals not participating. It is not clear if non-participants would be crowded out of the labor market or whether the returns are driven by increases in wages earned through entrepreneurship activities which would result in a net increase in employment.

While these caveats cannot be dismissed, the results presented here do provide the first experimental evidence about the effect of work experience on subsequent employment outcomes in a developing country. The effects are substantial, suggesting that short term training or employment programs that include work experience have transformative potential, and providing justification for further research on the topic.

2.7 References

Altonji, J.G. and Shakotko, R.A. (1987). "Do Wages Rise with Job Seniority?," Review of Economic Studies, vol. 54(3), pp. 437 – 459.

Altonji, J.G. and Williams, N. (2005). "Do wages rise with job seniority? A reassessment," Industrial and Labor Relations Review, Cornell University, vol. 58(3), pp. 370 – 397.

Beaman, L., Keleher, N. and Magruder, J. (2013). "Do Job Networks Disadvantage Women? Evidence from a recruitment experiment in Malawi," Mimeo.

Beaman, L. and Magruder, J. (2012). "Who gets the job referral? Evidence from a social networks experiment," American Economic Review, vol 102 (7) pp. 3574-3593, 2012

Betcherman, G., M. Godfrey, et al. (2007). "A Review of Interventions to Support Young Workers: Findings of the Youth Employment Inventory." World Bank Social Protection Discussion Paper 0715(October).

Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "The Determinants of Earnings: A Behavioral Approach." *Journal of Economic Literature* 39(4): 1137-1176.

Buchinsky, M., Fougere, D., Kramarz, F. and Tchernis, R. (2010). "Interfirm Mobility, Wages and the Returns to Seniority and Experience in the United States," Review of Economic Studies (2010) 77, 972–1001

Castel, V., Phiri, M. and Stampini, M. (2010). "Education and Employment in Malawi," African Development Bank Group, Working Paper Series, No. 110, June 2010.

Chirwa. E.W. and Matita, M.M. (2009). "The Rate of Return on Education in Malawi," Chancellor College Department of Economics Working Paper 2009/01.

Chirwa, E. W. and Zgovu, E. K. (2001) 'Does the Return to Schooling Depend on the Type of Employment? Evidence from the Rural Labour Market in Malawi.' Wadonda Consult Working Paper WC/03/01, February

Coleman, James S. "The Transition from School to Work." In *Research in Social Stratification and Mobility,* vol. 3, edited by Donald J. Treiman and Robert V. Robinson, pp. 27–59. Greenwich, CT: JAI Press, 1984

Devine, T. and Kiefer, N. (1991). "Empirical Labor Economics," Oxford: Oxford University Press.

Godlonton, S. (2013) "Employment Risk and Performance", University of Michigan Working Paper.

Granovetter, M. (1973). "The Strength of Weak Ties," American Journal of Sociology, vol. 78(6) pp. 1360-1380.

Heckman, J. J., R. J. LaLonde, et al. (1999). The economics and econometrics of active labormarket programs. Handbook of Labor Economics. O. Ashenfelter and D. Card, Elsevier. 3: 1865-2095.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24(3): 411-482.

Human Development Report. 2009.

IHS 2010/11. Integrated Household Survey. National Statistics Office of Malawi.

Jacob, Brian A. 2002. “Where the Boys Aren't: Non-cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review* 21: 589-598.

King, E.M., Montenegro, C.E. and Orazem, P.F. (2012). “Economic Freedom, Human Rights, and the Returns to Human Capital: An Evaluation of the Schultz Hypothesis”, Economic Development and Cultural Change, Vol. 61, No. 1 (October 2012), pp. 39-72

Kluve, J. (2006). “The Effectiveness of European Active Labor Market Policy.” IZA Discussion Paper No. 2018

Lam, D., Ardington, C. and Leibbrandt, M.(2011). “Schooling as a lottery: Racial differences in school advancement in urban South Africa”, Journal of Development Economics, Volume 95, Issue 2, July 2011, Pages 121-136, ISSN 0304-3878, 10.1016/j.jdeveco.2010.05.005.

Light, A. (1999). “High School Employment, High School Curriculum, and Post-School Wages.” *Economics of Education Review* 18 (May 1999): 291–309.

Light, A. (2001). “In-School Work Experience and the Returns to Schooling,” Journal of Labor Economics, Vol. 19, No. 1 (January 2001).

Light, A. and Ureta, M. (1995). “Early-Career Work Experience and Gender Wage Differentials,” Journal of Labor Economics, University of Chicago, vol. 13(1), pp. 121-54.

Lockheed, M.E., Verspoor, A.M. et al. (1991). “Improving Primary Education in Developing Countries,” New York, Oxford University Press.

Matita, M.M. and Chirwa, E.W. (2009). “The Impact of Education on Self-Employment, Farm Activities and Household Incomes in Malawi,” Chancellor College Department of Economics Working Paper 2009/02.

Meyer, Robert H., and Wise, David A. “High School Preparation and Early Labor Force Experience.” In *The Youth Labor Market Problem: Its Nature, Causes and Consequences,* edited by Richard B. Freeman and David A. Wise, pp. 277–344.

Chicago: University of Chicago Press for National Bureau of Economic Research, 1982.

Mincer, J. (1974). Schooling Experience and Earnings (New York: Columbia University Press.

Mwabu, G. and Schultz, T.P. (2000). "Wage Premiums for Education and Location fo South African Workers, by Gender and Race," Economic Development and Cultural Change, vol. 48(2), pp. 307-34.

Orazem, P.F. and King, E.M. (2008). "Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy," In Handbook of Development Economics, vol.4 ed. T, Paul Schultz and John Strauss. Amsterdam: North Holland.

Psacharopoulus, G. (1973). "Returns to Education: An International Comparison." Amstrerdam: Elsevier, San Francisco: Jossey-Bass.

Psacharopoulus, G. (1981). "Education, employment and inequality in LDCs," World Development, vol. 9(1), pp. 37-54.

Psacharopoulus, G. (1985). "Returns to Education: A Further International Update and Implications," Journal of Human Resources, vol. 20(4), pp. 583-604.

Psacharopoulus, G. (1994). "Returns to investment in education: A global update," World Development, vol.22(9), pp. 1325-1343.

Pugatch, T. (2012). "Bumpy Rides: School to Work Transitions in South Africa" IZA Discussion Paper No. 6305.

Ruhm, Christopher. (1995). "The Extent and Consequences of High School Employment." *Journal of Labor Research* 16 (Summer 1995): 293–303.

Ruhm, C. (1997). "Is High School Employment Consumption or Investment?" *Journal of Labor Economics* 15 (October 1997): 735–76

Spence, M. (1973). "Job Market Signalling," Quarterly Journal of Economics, vol. 87(3) pp.335 – 374.

Topel, R.H. (1991). "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," Journal of Political Economy, vol. 99(1), pp. 145 – 176.

World Development Report: Jobs (2013). World Bank.

Figure 2.1: Timeline of experiment, and data collection activities

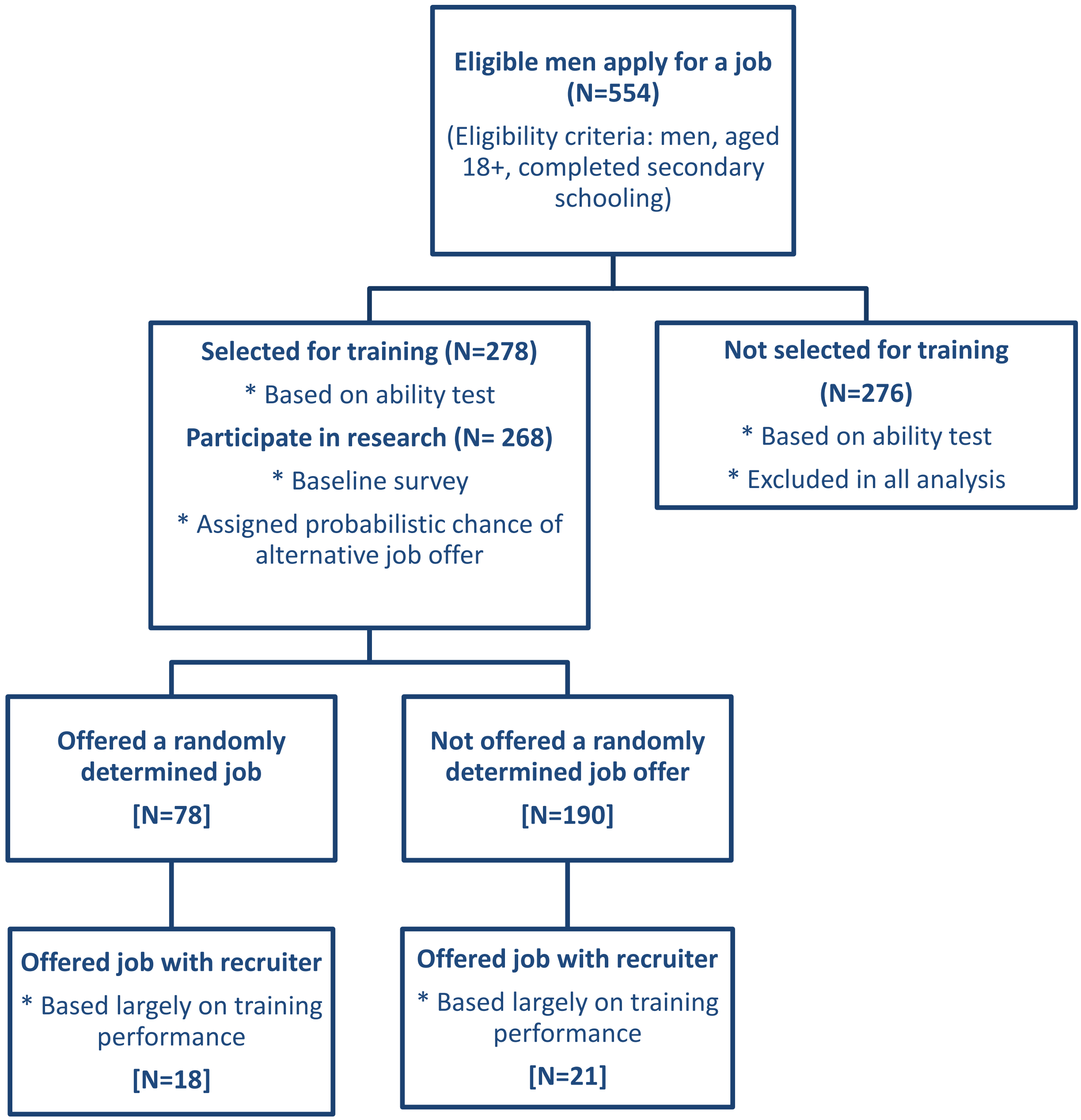


Figure 2.2: Estimated employment impact of job offer by month (IV estimates)

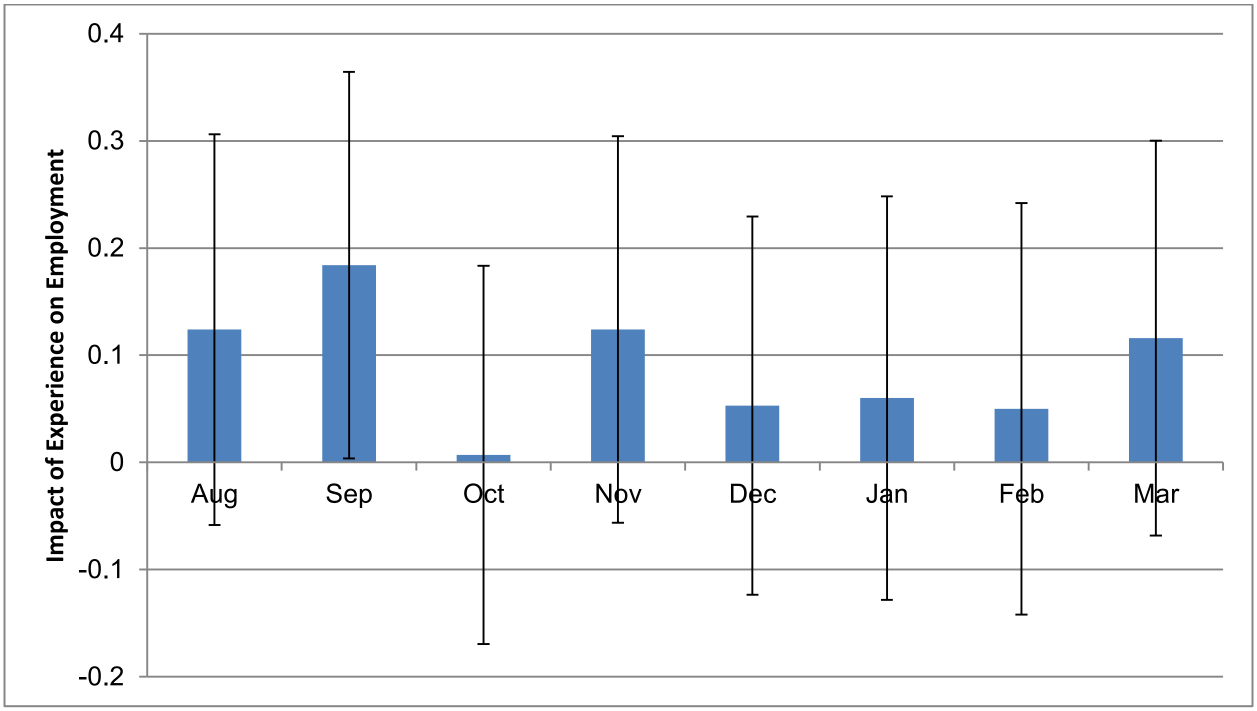


Figure 2.3: Estimated wage impact of job offer by month (IV estimates)

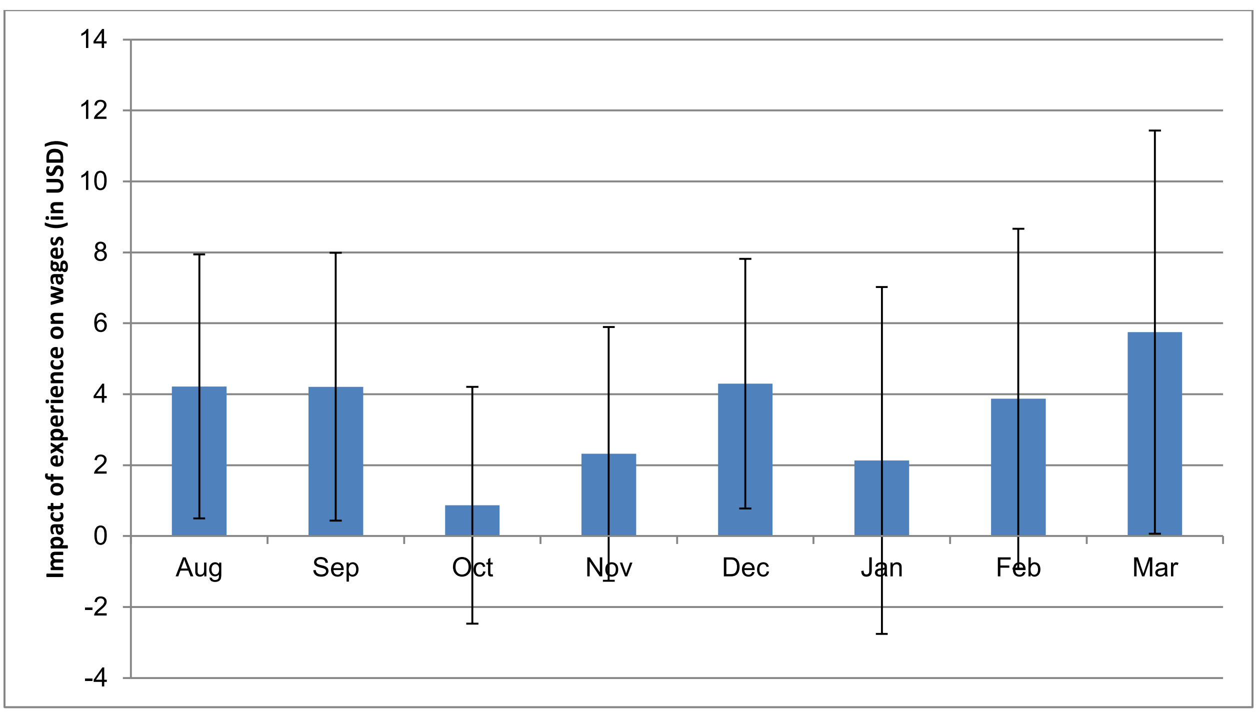


Figure 2.4: Distribution of wages

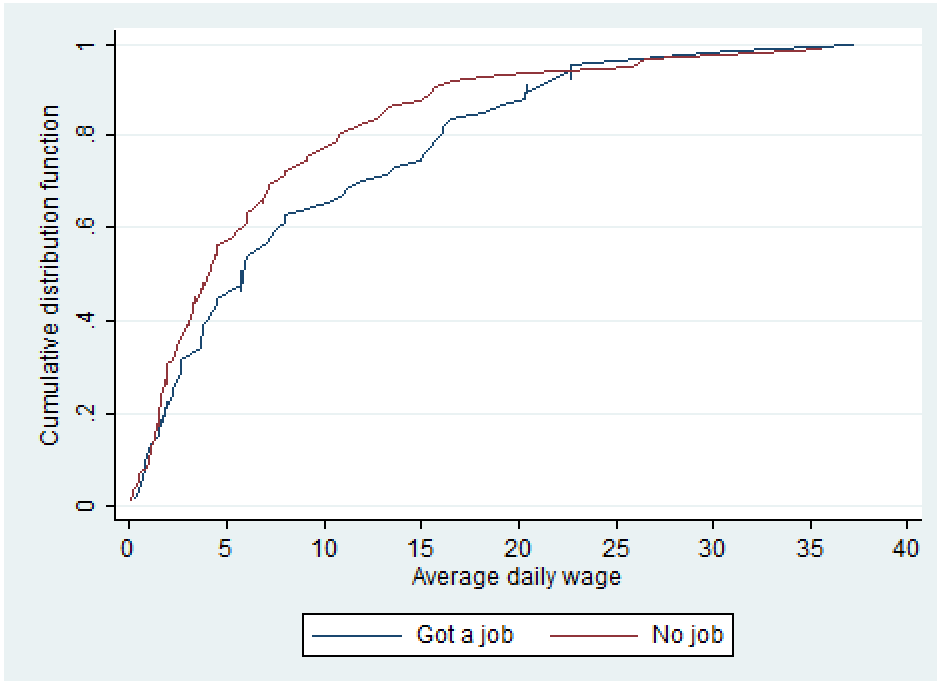


Figure 2.5: Estimated employment impact by ability of job offer by month (IV estimates)

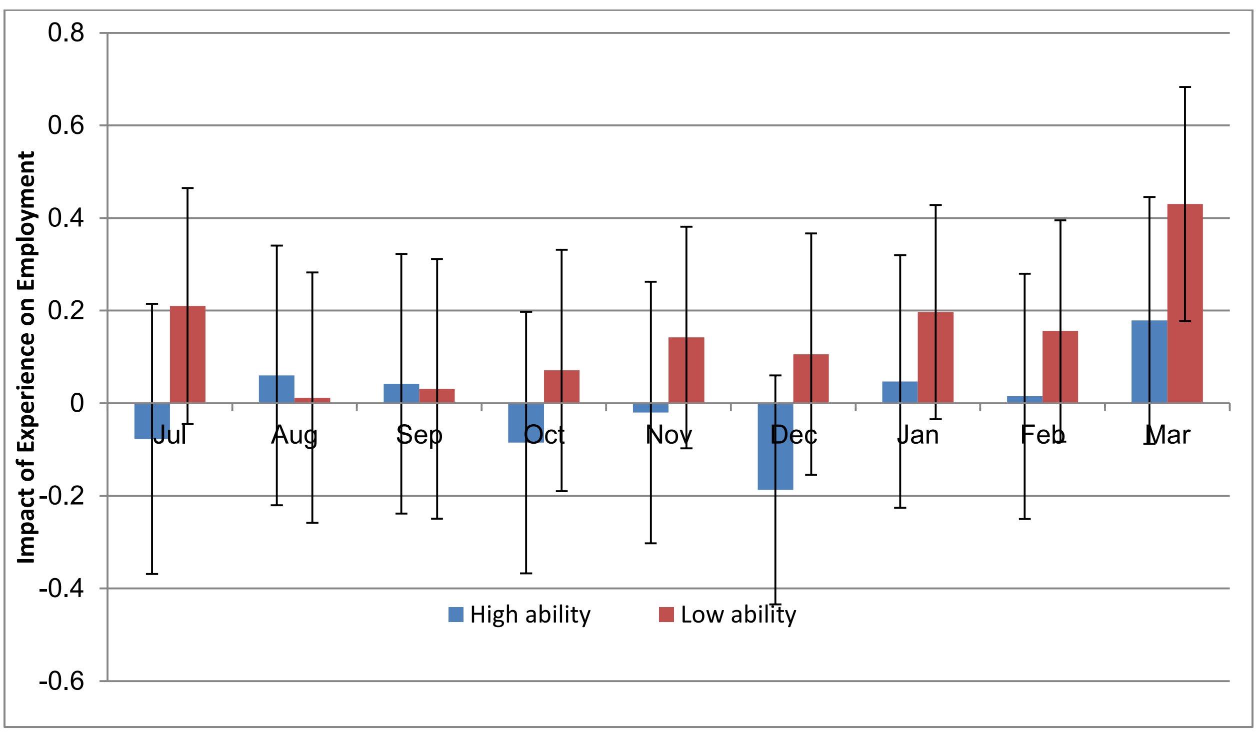


Table 2.1: Sample size and attrition

Treatment conditions:	N (1)	Mean (2)	SD (3)
0% Probability	53	0.811	0.395
1% Probability	56	0.857	0.353
5% Probability	52	0.827	0.382
50% Probability	54	0.852	0.359
75% Probability	28	0.929	0.262
100% Probability	25	0.840	0.374
Full sample:	**268**	**0.847**	**0.361**
p-value of F-test of joint significance:			
0% = 1% = 5% = 50% = 75% = 100%		0.827	

p-values of t-tests of pair-wise differences:

	1%	**5%**	**50%**	**75%**	**100%**
0%	0.510	0.826	0.564	0.168	0.745
1%		0.666	0.939	0.396	0.844
5%			0.724	0.233	0.882
50%				0.364	0.893
75%					0.376

Notes:

Individuals were assigned to one of the six treatment groups. If they received a 0-percent chance of an alternative (i.e. in 0% probability treatment group) then they had no chance of receiving the alternative job. If they were assigned to the 1% probability group then they had 1 percent chance of receiving an alternative job. Similarly for the 5-, 50-, 75- and 100 percent probability groups. There were twice as many assigned to the high probability groups as compared to the lower groups due to budgetary considerations. The p-values denote the p-value associated with the F-test of whether the mean finding rate is the same in all treatment groups or in the case of the table the pair-wise t-test of differential attirion rates.

Table 2.2: Sample and Attrition

	Baseline N=268		Follow-Up N=227		Difference	
	Mean	SD	Mean	SD	(3) - (1)	
	(1)	(2)	(3)	(4)	(5)	
Demographics:						
Age	25.604	4.638	25.718	4.662	-0.114	
Married	0.172	0.378	0.172	0.378	0.000	
Any child?	0.164	0.371	0.167	0.374	-0.003	
Number of children	0.299	0.784	0.313	0.811	-0.014	
Number of fin dependents	7.959	9.355	8.264	9.406	-0.305	
Years of education	13.183	0.940	13.220	0.938	-0.037	
Income (USD, 3 months)	206.123	228.803	210.617	237.777	-4.494	
Ability score	-0.001	1.003	0.030	1.017	-0.031	
Tribe:						
Chewa	0.310	0.463	0.300	0.459	0.010	
Lomwe	0.108	0.311	0.110	0.314	-0.002	
Ngoni	0.164	0.371	0.181	0.386	-0.016	**
Tumbuka	0.190	0.393	0.189	0.393	0.001	
Other	0.201	0.402	0.198	0.400	0.003	
Education and Work:						
Ever worked?	0.869	0.338	0.863	0.344	0.006	
Ever worked with recruiter?	0.104	0.306	0.097	0.296	0.008	
Any work in last month	0.646	0.479	0.665	0.473	-0.020	
Any work in last 6 months	0.869	0.338	0.890	0.314	-0.020	*
Frac of 6 mths worked	2.657	2.176	2.727	2.175	-0.070	
Any job search last month	0.116	0.320	0.110	0.314	0.006	

Notes:
The baseline sample consists of 268 individuals who participated in the recruitment process and experiment discussed in Section 2. The follow-up sample (227 respondents) is the main sample used in this paper. The ability score is determined prior to the experiment. It consists of a numeracy and literacy component, and has been standardized.

Table 2.3: First Stage: Using dummy indicators for each treatment group to predict any job offer (recruiter or random job)

Dependent Variable:	**Job offer or recruiter's job offer**		
	(1)	**(2)**	**(3)**
1% Job Guarantee	0.025	0.030	-0.004
	[0.081]	[0.078]	[0.083]
5% Job Guarantee	0.047	0.045	0.038
	[0.085]	[0.079]	[0.085]
50% Job Guarantee	0.402***	0.423***	0.439***
	[0.094]	[0.090]	[0.093]
75% Job Guarantee	0.568***	0.543***	0.565***
	[0.105]	[0.104]	[0.108]
100% Job Guarantee	0.837***	0.860***	0.866***
	[0.057]	[0.055]	[0.067]
Constant	0.163***	0.804***	0.544
	[0.057]	[0.153]	[0.370]
Observations	227	227	227
R-squared	0.327	0.382	0.431
Stratification cell FE's	No	Yes	Yes
F-stat (of instruments)	101.11	87.47	76.79
Average of dep variable		0.361	

Notes:

The sample used here is the sample of 227 men found at follow-up. The zero percent chance of alternative employment treatment group is the omitted category in these regressions. The dependent variable "Got a job" is whether or not the individual received an alternative job offer. Stratification cell fixed effects are included as the randomization was conducted by stratifying on baseline ability and whether the individual had ever worked with the recruiter previously. The set of covariates includes: age, marital status, education dummies, a dummy indicator for whether the respondent has any children, the number of children that the respondent has, ability score (a composite measure of numeracy and literacy scores), dummy indicators for tribe, a dummy indicator if the respondent has any work experience, reports any work in the past month and any job search in the past month, and the number of months in the last six months he has worked.*** denotes statistical significance at the 1 percent level, ** 5 percent level, and * 1 percent level. Robust standard errors are reported.

Table 2.4: Returns to Work Experience: Extensive Margin

Dependent Variable:	**Frac. months employed**			**Frac. months looked for work**			**Ave # concurrent jobs**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Got a job or recruiters job	0.106	0.128	0.139*	0.084	0.105	0.113	0.079	0.098	0.071
offer (IV)	[0.086]	[0.086]	[0.076]	[0.091]	[0.090]	[0.079]	[0.082]	[0.081]	[0.072]
Constant	0.376***	0.538***	-0.015	0.395***	0.520***	0.043	0.597***	0.754***	0.024
	[0.041]	[0.128]	[0.349]	[0.043]	[0.140]	[0.355]	[0.038]	[0.114]	[0.275]
Stratification cell FE's	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Other covariates?	No	No	Yes	No	No	Yes	No	No	Yes
Observations	227	227	227	227	227	227	227	227	227
R-squared		0.087	0.282		0.085	0.279	0.019	0.047	0.249
Ave of dep variable (no job)		0.421			0.586			0.532	

Notes:

The regressions are IV estimates, where dummy indicators for the treatment assignment (i.e. assignment to a 0-, 1-, 5-, 50-, 75-, or 100-percent chance of employment) are used to instrument for the binary indicator got a job offer from recruiter or through random determination.

The fraction months employed variable is calculated as the number of months the individual was employed over the last 8 months, divided by 8. Similarly, the fraction months looked for work variable is computed using a retrospective calendar history, and is calculated as the number of months the individual actively sought work over the last 8 months, divided by 8.Lastly, the average number of concurrent jobs is the average of the total number of jobs held each month across the 8 month period.

Stratification cell fixed effects are included as the randomization was conducted by stratifying on baseline ability and whether the individual had ever worked with the recruiter previously.The set of covariates includes: age, marital status, education dummies, a dummy indicator for whether the respondent has any children, the number of children that the respondent has, ability score (a composite measure of numeracy and literacy scores), dummy indicators for tribe, a dummy indicator if the respondent has any work experience, reports any work in the past month and any job search in the past month, and the number of months in the last six months he has worked.*** denotes statistical significance at the 1 percent level, ** 5 percent level, and * 1 percent level. Robust standard errors are reported.

Table 2.5: Returns to Work Experience: Intensive Margin												
Dependent Variable:	**Ave hrs worked per week**			**Hourly wage**			**Ave daily wage (incl. Unemployed)**			**Log (Ave daily wage)**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
Got a job or recruiters job offer	-0.576	-0.567	-0.691*	5.463	5.836	7.854*	3.801*	4.191*	3.928**	0.668*	0.687*	0.605*
	[0.381]	[0.376]	[0.387]	[4.361]	[4.404]	[4.173]	[2.149]	[2.218]	[1.885]	[0.373]	[0.387]	[0.354]
Constant	3.223***	4.033***	4.720***	21.559***	14.082***	18.413	4.133***	10.784	-1.611	1.206***	0.621	0.216
	[0.185]	[0.433]	[1.139]	[2.112]	[4.321]	[15.205]	[0.864]	[7.161]	[5.779]	[0.173]	[0.559]	[1.135]
Stratification cell fixed effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Other covariates?	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	166	166	166	167	167	167	227	227	227	164	164	164
R-squared	0.029	0.069	0.154		0.035	0.199		0.045	0.251		0.036	0.262
Ave of dep variable (no job)		23.265			0.489			5.079			1.361	

Notes:

The regressions are IV estimates, where dummy indicators for the treatment assignment (i.e. assignment to a 0-, 1-, 5-, 50-, 75-, or 100-percent chance of employment) are used to instrument for the binary indicator got a job offer from recruiter or through random determination.

The average daily wage is calculated using the restrospective job work history. The average daily wage is calculated as the average wage on the individual's main job in the last month. Columns 1 through 3, those who are unemployed are coded as 0's. Columns 4 through 6 uses the logged wage, therefore for individuals who earned $0 across all eight months are omitted.

Stratification cell fixed effects are included as the randomization was conducted by stratifying on baseline ability and whether the individual had ever worked with the recruiter previously.The set of covariates includes: age, marital status, education dummies, a dummy indicator for whether the respondent has any children, the number of children that the respondent has, ability score (a composite measure of numeracy and literacy scores), dummy indicators for tribe, a dummy indicator if the respondent has any work experience, reports any work in the past month and any job search in the past month, and the number of months in the last six months he has worked.*** denotes statistical significance at the 1 percent level, ** 5 percent level, and * 1 percent level. Robust standard errors are reported.

	Any job held in past 8 months:			Num months in each occupation in past 8 months:			Modal occupation in past 8 months		
	Avg dep var (no job)	Coeff	SE	Avg dep var (no job)	Coeff	SE	Avg dep var (no job)	Coeff	SE
Occupation:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Professional, technical, and related workers	0.368	0.158	[0.108]	1.515	0.661	[0.599]	0.475	-0.075	[0.142]
Administrative and managerial workers	0.007	0.052	[0.040]	0.007	0.279	[0.251]	0.000	0.083	[0.053]
Clerical and related workers	0.213	0.150	[0.106]	0.691	0.695	[0.485]	0.212	0.057	[0.125]
Sales workers	0.044	-0.015	[0.043]	0.096	0.117	[0.190]	0.030	0.009	[0.055]
Service workers	0.066	-0.053	[0.040]	0.419	-0.352	[0.259]	0.091	-0.056	[0.057]
Agriculture, animal husbandry, and forestry workers, fishermen, and hunters	0.066	-0.032	[0.038]	0.346	-0.130	[0.227]	0.081	-0.019	[0.052]
Production and related workers, transport equipment operators, and labourers	0.110	-0.029	[0.061]	0.471	-0.074	[0.284]	0.111	0.001	[0.079]

Table 2.6: Shifts in occupations

Notes:

The regressions are IV estimates, where dummy indicators for the treatment assignment (i.e. assignment to a 0-, 1-, 5-, 50-, 75-, or 100-percent chance of employment) are used to instrument for the binary indicator got a job offer from recruiter or through random determination.

Stratification cell fixed effects are included as the randomization was conducted by stratifying on baseline ability and whether the individual had ever worked with the recruiter previously.The set of covariates includes: age, marital status, education dummies, a dummy indicator for whether the respondent has any children, the number of children that the respondent has, ability score (a composite measure of numeracy and literacy scores), dummy indicators for tribe, a dummy indicator if the respondent has any work experience, reports any work in the past month and any job search in the past month, and the number of months in the last six months he has worked.*** denotes statistical significance at the 1 percent level, ** 5 percent level, and * 1 percent level. Robust standard errors are reported.

Table 2.7: Contract type			
Dependent Variable:	**Unit of pay (1 = daily, 2 = weekly, 3 = fortnightly; 4 = monthly)**		
	(1)	**(2)**	**(3)**
Got a job or recruiters job offer (IV)	-0.576	-0.567	-0.691*
	[0.381]	[0.376]	[0.387]
Constant	3.223***	4.033***	4.720***
	[0.185]	[0.433]	[1.139]
Stratification cell FE's	No	Yes	Yes
Other covariates?	No	No	Yes
Observations	166	166	166
R-squared	0.029	0.069	0.154
Ave of dep variable (no job)		3.169	

<u>Notes:</u>

The regressions are IV estimates, where dummy indicators for the treatment assignment (i.e. assignment to a 0-, 1-, 5-, 50-, 75-, or 100-percent chance of employment) are used to instrument for

The average daily wage is calculated using the restrospective job work history. The average daily wage is calculated as the average wage on the individual's main job in the last month. Columns 1 through 3, those who are unemployed are coded as 0's. Columns 4 through 6 uses the logged wage, therefore for individuals who earned $0 across all eight months are omitted. Ave hours worked per

Stratification cell fixed effects are included as the randomization was conducted by stratifying on baseline ability and whether the individual had ever worked with the recruiter previously.The set of covariates includes: age, marital status, education dummies, a dummy indicator for whether the respondent has any children, the number of children that the respondent has, ability score (a composite measure of numeracy and literacy scores), dummy indicators for tribe, a dummy indicator if the respondent has any work experience, reports any work in the past month and any job search in the past

Table 2.8: Channels

Panel A: Social Networks:	Avg dep var (1)	Coeff (2)	SE (3)
Heard about a job opportunity	0.438	0.234**	[0.115]
# job opportunities	0.795	0.126	[0.268]
Secured a job opportunity	0.091	0.085	[0.056]
# job opportunities secured	0.080	0.077	[0.056]
Panel B: Signalling:			
Used any reference letter for a job in last 8 months	0.648	-0.074	[0.110]
Panel C: Wage Expectations:			
Self-reported monthly reservation wage	361.873	121.253	[84.563]

Notes:

The regressions are IV estimates, where dummy indicators for the treatment assignment (i.e. assignment to a 0-, 1-, 5-, 50-, 75-, or 100-percent chance of employment) are used to instrument for the binary indicator got a job offer from recruiter or through random determination.
Stratification cell fixed effects are included as the randomization was conducted by stratifying on baseline ability and whether the individual had ever worked with the recruiter previously.The set of covariates includes: age, marital status, education dummies, a dummy indicator for whether the respondent has any children, the number of children that the respondent has, ability score (a composite measure of numeracy and literacy scores), dummy indicators for tribe, a dummy indicator if the respondent has any work experience, reports any work in the past month and any job search in the past month, and the number of months in the last six months he has worked.*** denotes statistical significance at the 1 percent level, ** 5 percent level, and * 1 percent level. Robust standard errors are reported

Table 2.9: Heterogeneity of wage and employment impacts			
Panel A: Ability Inteactions	**Frac. months worked**	**Avg daily wage (incl. unemployed)**	**Logged (Avg daily wage)**
	(1)	**(2)**	**(3)**
Got a job	0.119	4.443**	6.711**
	[0.074]	[1.820]	[3.137]
Ability score X Got job	-0.169**	-3.046*	-5.654*
	[0.079]	[1.773]	[3.038]
Ability score	0.099	-1.396	2.636
	[0.099]	[1.741]	[4.166]
Panel B: Degree interactions			
	(1)	**(2)**	**(3)**
Got a job	0.057	7.431*	1.545**
	[0.209]	[4.407]	[0.759]
Degree X Got a job	0.359	-21.501	-4.144
	[1.254]	[24.915]	[3.204]
Degree	-0.033	0.000	3.513**
	[0.000]	[12.706]	[1.645]
Panel C: Current labor attachment interactions:			
	(1)	**(2)**	**(3)**
Got a job	0.761	6.851	5.126
	[1.108]	[23.227]	[4.178]
Any work in last month X Got a job	-0.918	-4.075	-6.233
	[1.532]	[33.295]	[5.784]
Any work in last month	0.334	2.087	2.295
	[0.483]	[10.858]	[1.778]
Panel D: Any previous experience interactions:			
	(1)	**(2)**	**(3)**
Got a job	-0.233	9.528	8.844
	[0.478]	[9.609]	[9.874]
Ever worked X Got a job	0.530	-8.900	-11.214
	[0.742]	[15.287]	[13.496]
Ever worked	-0.189	4.898	4.136
	[0.277]	[5.617]	[4.853]

Notes:

The probability of alternative employment (P_i) and the interaction of the baseline characteristic and the probability of alternative employment assigned ($Base_i * P_i$) are used to instrument for the binary indicator JO_i and the interaction of the baseline characteristic and the job offer ($Base_i * JO_i$). The fraction months employed variable is calculated as the number of months the individual was employed over the last 8 months, divided by 8. The average daily wage is calculated using the restrospective job work history. Stratification cell fixed effects are included as the randomization was conducted by stratifying on baseline ability and whether the individual had ever worked with the recruiter previously.The set of covariates includes: age, marital status, education dummies, a dummy indicator for whether the respondent has any children, the number of children that the respondent has, ability score (a composite measure of numeracy and literacy scores), dummy indicators for tribe, a dummy indicator if the respondent has any work experience, reports any work in the past month and any job search in the past month, and the number of months in the last six months he has worked.*** denotes statistical significance at the 1 percent level, ** 5 percent level, and * 10 percent level. Robust standard errors are reported.

Chapter 3

Employee Crime and Monitoring

3.1 Introduction

Employee crime is costly and widespread. Snyder (1991) estimates the losses incurred from employee theft, excluding fraudulent reporting of financial statements, to be approximately $120 billion per annum in the United States alone. Statistics on the extent of employee crime are less readily available in developing countries, but the available evidence indicates high levels of employee crime (PwC, 2002; Chandra et al., 2001; Fafchamps, 2004; UN, 2003). In addition to the costs to individual businesses, corruption broadly construed has been highlighted as a constraint to growth (Schleifer and Vishny, 1993; Mauro, 1994). Also, small scale firms also cite employee theft as a reason not to hire workers (Fafchamps, 2004). Therefore, employee crime may thwart firm growth and job creation. Understanding strategies that may reduce employee crime is important. In this paper I examine the role of one strategy – auditing of monetary transactions – and the impact of this strategy on employee crime.

Whether and to what extent auditing will affect employee crime is theoretically ambiguous. Becker's seminal work (1968) examines the optimal allocation of resources to the detection and punishment of crime. He shows that the optimal amount of enforcement depends on the cost of detecting and punishing individuals, as well as the behavior change of offenders to enforcement. Becker and Stigler (1974) apply this model to employee crime and show that both higher wages and the positive probability of audit deter corruption. The economic rationale is clear. The higher the cost of committing a crime, conditional on being caught, the lower the expected payoff to committing that crime. Behavioral economists and psychologists have challenged this rational-expectations based intuition, however. Frey (1993a and 1993b) suggest that implicit psychological contracts between an employer and his employee could be undermined in the presence of a non-zero audit probability. Specifically, increased monitoring may in fact increase employee crime or lower effort. Chang and Lai (1999) combine the rational actor model with the behavioral insights. From this model they derive theoretically ambiguous predictions for how employee crime responds to monitoring intensity in the workplace. However, their model does not provide conditions under which either economic incentives or psychological disincentives dominate.

The existing empirical literature finds mixed results about the effect of monitoring on employee crime. An extensive review of the literature examining how employees respond to incentives in the workforce broadly is provided in Prendergast (1996). This review largely finds that individuals act in accordance with a rational cheater model. Specifically, individuals respond to incentives induced by monitoring by changing their behavior when the marginal benefit of the alternative outweighs its marginal cost.

Although theoretical models identify two key channels to influence employee behavior – wages and monitoring – the empirical literature has largely focused on the impact of wages on employee effort (Laezear 1996; Paarsch and Shearer 1996). A handful of papers have examined the trade-off between wages and supervision (Groshen and Krueger 1990; Kruse 1992; Rebitzer 1995; and di Tella and Schargodsky 2003). To my knowledge, there is only one paper that focuses explicitly on the impact of monitoring on employee effort, specifically opportunistic behavior (Nagin et al., 2002). Nagin et al. (2002) use experimentally induced variation in the audit rate at a call center firm in the United States to examine the impact of auditing on employee effort and opportunistic behavior. The probability of an audit ranges from zero to 15 percent. Employees exhibit heterogeneous responses to monitoring: while some employees reduced their effort when the audit rate declined, a substantial fraction did not. Moreover, some employees reduced their effort when the audit rate increased. These mixed empirical results from one study from one firm in the United States suggest that there is scope to learn a great deal more about how employees respond to changes in the degree of supervision in the workplace.

A key limitation when examining the impact of auditing on worker effort or specifically employee crime is that auditing is often endogenous to the prevalence and cost of employee crime. Within firm or across firm variation in audit rates are likely to arise in part due to differences in the existence or extent of crime. Specifically, increases in auditing may result precisely because there is a problem with employee effort or crime. Moreover, it is likely that employees who are suspected of being likely to engage in employee crime are also those likely to be supervised with greater intensity. This makes it difficult to ascertain the causal impact of changes in auditing on employee crime.

A second limitation arises due to difficulties in measuring the audit rate. Many studies that examine the impact of monitoring use the ratio of supervisors to staff rather than an actual measure of monitoring by those supervisors (Groshen and Krueger, 1990). Because most supervisors only allocate some of their time to monitoring, much of the true variation in monitoring is not captured with this measure. Also, under this definition the monitoring rate is likely correlated with other human resource protocols that independently affect employee effort (and specifically employee crime), therefore introducing omitted variable bias.

A third challenge relates to the measurement of employee effort and employee crime. Many studies use measures of worker productivity; for some industries this is relatively easy, for example if piece-rate wages are paid (Laezear 1996; Paarsch and Shearer 1996). However, in many other cases productivity is not easily observed. Employee crime is also difficult to observe. Many studies that specifically focus on employee crime utilize susceptibility to crime rather than actual employee transgressions due to the lack of availability of such data (e.g. Nagin and Paternoster, 1993; Barnes and Lambell, 2007). Studies that measure crime directly often use an aggregated measure. For example, Rickman and Witt (2007) use actual record employee theft aggregated by region due to data availability.

This paper exploits experimentally induced variation in the degree of supervision. Specifically, I measure how interviewers employed in a short term contracting market in Malawi alter their opportunistic behavior in response to changes in the audit rate. Like the Nagin et al. (2002) paper, the approach taken here overcomes the standard challenges to identification and measurement. In this paper the monitoring rate is exogenously induced; the measurement of monitoring is well-defined and accurate; and the measurement of employee crime (although one-dimensional) is interviewer-specific and captures actual rather than perceived behavior. This paper is most similar to the Nagin et al. (2002) paper but makes a number of new contributions. First, I observe variation in monitoring over a wider spread of possible audit rates; it provides the first evidence in a developing country setting in a more informal labor market.; I examine employee theft rather than opportunistic behavior use actual theft; and I exploit within employee variation rather than across employee variation to measure the responses. Moreover, the results in Nagin et al. (2002) are mixed and more evidence on behavioral responses to varying monitoring regimes is needed.

The specific short term contracting market is the market for interviewers. Employees in this context are interviewers who are required to handle cash transactions as part of their data collection efforts. Each work day interviewers are provided a set amount of stock and cash. At the end of each day the remainder of the stock and cash are collected. The amount of stock and cash can be reconciled with the questionnaires submitted. Whether or not interviewers are audited and the probability that an individual interviewer is audited on a given day is randomized daily. Each day all employees face the same chance of being audited; for example, on a day with a 50 percent audit probability, half of the interviewers will be randomly selected for audit. This variation identifies the impact of monitoring on two measures of employee crime: the probability of stealing cash, and the amount of cash stolen. The data are collected from two distinct rounds of data collection with slightly different protocols. The data comprises 34 distinct employees and 36 different days for a total of 499 observations. Different employees work in the different rounds; and even within round not all employees work all days. Also, in round two, not all individuals are audited each day reducing the number of interviewer-day observations.

I find that the prevalence of employee crime decreases as the monitoring rate increases. In the first round of data collection, the probability of stealing any money declines by between 16.7 and 21.7 percentage points when moving from no monitoring to full monitoring. The response is considerably stronger in round two, where I find interviewers were approximately 55 percentage points less likely to pocket any money in the case of no monitoring compared to full monitoring. The process in round two was more transparent to employees, and the audit rate was more credible. In both cases, the effect of monitoring is statistically significant, and is robust to the inclusion of interviewer and day-of-week fixed effects.

Increasing the monitoring rate from no monitoring to full monitoring reduced the amount of money stolen per person by five to nine kwacha. While in dollar terms this is a small amount of money, in percentage terms this is a large reduction ranging from a 50 to 80 percent reduction depending on the round and empirical specification used. The estimated effect is not statistically significant at the 10 percent level in either individual round due to limited statistical power. However, when using the pooled data the reduction in the amount stolen is statistically significant at the 5 percent level.

The paper proceeds as follows. Section 3.2 sets out a simple conceptual framework of employee crime. Section 3.3 provides relevant details about the specific context, the experiment implemented, the data used in this paper and the measurement of the audit rate and employee crime. Section 3.4 presents the empirical strategy and Section 3.5 presents and discusses the results. Section 3.6 concludes.

3.2 Employee crime – Monitoring and fines

The decision to engage in corrupt activities is a decision made under uncertainty. That uncertainty operates through two key channels. First, there is uncertainty as to whether the interviewer will be caught. Second, there is uncertainty with respect to the penalty if caught. The basic model set out below focuses on an interviewer's decision to commit employee theft.

There are two states of the world, one in which corruption is detected and another in which it is not. Interviewers face a standard concave utility function such that $U' > 0$ and $U'' < 0$. The employee's choice variable is the extent of participation in corrupt activities d – here, measured as the amount of money stolen. The present discounted value of future income is captured by A. Conceptually, this income is a composite of potential income from the current employer and other employers. The probability of being hired in the future (either by the current employer or other employers) is denoted by q. The probability that an employee who has stolen money is caught is equivalent to the percent of interviewers monitored on any particular day, p. In other words, if an employee cheats and he is audited he is caught. If an interviewer is caught stealing he faces a penalty, $F(d)$. The penalty imposed translates into a reduction in the interviewers' probability of accessing future employment opportunities, and as such results in a lower present discounted future income in the state of the world in which he is caught. Employees are not immediately fired as there are large frictions for the employer to replace the employee, including search and training costs. Therefore, it is inefficient to summarily dismiss any employee caught cheating. The penalty imposed can be thought of as a negative shock to reputation with both the current and other employers. This penalty is an increasing function of the amount stolen (i.e. $F'(d)>0$); those who are caught stealing more suffer a larger cost to their reputation and therefore a lower probability of future job opportunities. Assume $F''(d)<0$, i.e. the penalty is increasing at a decreasing rate in the level of employee crime. In this case the expected utility function is:

$$EU = (1-p)U[d+qA] + pU[d+(q-F(d))A] \tag{1}$$

The first-order condition for an interior maximum is:

$$(1-p)U'[d+qA] + pU'[d+\big(d-F(d)\big)A](1-F') = 0 \tag{2}$$

The second order condition for a maximum is:

$$SOC = (1-p)U'[d+qA] + pU'\big[d+\big(d-F(d)\big)A\big](1-F')^2 \leq 0 \tag{3}$$

The first comparative static of interest is how corrupt behavior d depends on the monitoring rate p. Differentiating (2) and simplifying yields:

$$\frac{\partial d}{\partial p} = \frac{1}{SOC}\big[U'[d+qA] + U'[d+\big(d-F(d)\big)A](1-F')\big] < 0 \tag{4}$$

The above derivative is negative because marginal utility is positive in both states of the world; the SOC is positive and π_d is a probability. As the monitoring rate increases (i.e. risk of detection rises) the likelihood or amount of employee crime committed should decline. Including a nominal fixed cost of being caught that embodies distaste for stealing, or the stigma of being caught will ensure that the optimal amount stolen under full monitoring is not zero.

3.3 Background and intervention

3.3.1 Context

Nationally representative firm level data providing employee crime prevalence is not available in Malawi. However, a survey of small traders conducted in 1999 and 2000 shows that 33 percent of these traders were victims of inventory theft in the past 12 months. The reported maximum value of goods stolen relative to annual sales was as high as 42.1 percent. Only nine percent of the traders suspected an employee in the event of inventory theft but an additional 11 percent of traders reported not hiring workers for fear of theft (Fafchamps, 2004). Interestingly, the National Bank of Malawi promotes their VISA debit cards by citing reduced employee theft as a benefit (National Bank of Malawi, 2008). Both the statistical evidence and the market response suggest that employee crime is an important challenge to firms and small traders in Malawi.

I conduct this experiment in the context of a market for short term skilled labor contracts. This market is characterized by fairly well educated employees who are hired shortly before the start of a contract to work for fixed periods of employment. Interviewers often work on overlapping contracts – before the one contract terminates they find a new contract and switch employers. Mostly, neither the employer nor the employee is required to give notice.

In this paper, employees are hired as interviewers. Their responsibilities include interviewing a predefined sample of men across a period of several weeks. Interviewers were required make a cash payment to each respondent and then offer the respondent the opportunity to purchase reduced-priced condoms. There are multiple forms of employee crime that can arise in this setting. First, employees can steal money that remains at the end of the work day. Second, employees can pay respondents too little money and pocket the difference. Third, employees can complete ghost questionnaires and pocket the money intended for a respondent. Optimally, one would like to measure all three directly. Unfortunately, good data is only available on the first type of crime and is therefore the focus of this paper.

Evidence regarding the other types of employee crime suggests that it is very low. For example, the number of interviews conducted does not vary with the audit rate. Although I cannot directly measure the prevalence of ghost questionnaires there is suggestive evidence that the general prevalence of this type of dishonest behavior is low and not differential by the monitoring rate. First, supervisors of the interviewers perform spot checks to ensure that interviewers actually visited respondents. From these spot checks only one case of an interviewer paying the respondent too little was observed, and no ghost respondents were noted. Second, the survey data collected by the interviewers was part of a panel study; using information from one of the other waves of data collection I construct a measure of inconsistencies in time-invariant variables by day of interview across survey wave. A higher frequency of inconsistencies between the two waves of data would suggest a higher probability that the interviewer did not interview the correct respondent or any respondent. Importantly, this measure of inconsistency does not vary by the audit rate on the day of the interview.

3.3.2 Monitoring experiment

The data used in this paper is pooled across two rounds of experimentation. In this section I outline the details of each round and note the key differences between the two rounds.

3.3.2.1. Round 1:

The detailed timeline of a typical work day for the subjects of this experiment is presented in Figure 3.1. The first round of data collection includes 13 distinct interviewers working on 19 different days. Not all interviewers worked each day. There are a total of 193 interviewer-day level observations from this round of data. In this initial round of data collection the monitoring rate was not explicitly randomized but determined in an ad-hoc manner before each work day. It was determined by the experimenter but prior to examining the level of theft on the previous day. To minimize interference with the project for which the interviewers were hired the monitoring rate was not altered every day during this round of data collection. There were three different monitoring rates: 0 percent, 33 percent and 100 percent. The audit rate was communicated directly to the field supervisors, who were then required to communicate this to all interviewers at the start of each work day. On days where the audit rate was 33 percent, the identity of the interviewers to be audited was not revealed to supervisors in advance. This was enforced to minimize the extent to which supervisors could collude with interviewers.

At the beginning of each work day interviewers received 300 Malawian kwacha; a set number of condoms; and questionnaires. Supervisors communicated the monitoring rate by announcing how many of the interviewers would be audited at the end of the work day. During the work day each interviewer was required to give each respondent 30 Malawian Kwacha and then offer the respondent the opportunity to purchase reduced-priced condoms. Condoms were sold at the following rates: one condom for two kwacha or a pack of three condoms for five kwacha. Respondents could buy up to 30 kwacha worth of condoms. Interviewers typically completed four interviews and sold an average of 7.29 condoms per day (Table 3.1). At the end of each work day interviewers were required to submit all questionnaires to their supervisor and all remaining condoms and coins to the experimenter.

The experimenter collected data about transaction accuracy for all interviewers every day (effectively, implementing a 100 percent audit rate). This process involved carefully reconciling the amount of money returned and comparing it to the amount that should have been returned given the number of questionnaires completed and condoms sold. A list of all interviewers for which there was a discrepancy and the amount of the discrepancy was communicated to the supervisors. On days when the full-monitoring regime had been announced to interviewers, supervisors privately communicated discrepancies to the interviewers for whom cheating was detected. When there was partial monitoring supervisors spoke to the subset of interviewers who were randomly chosen to be subject to auditing, and on days on which there was no monitoring no-one was informed about whether they had returned too little money. Other than communicating the discrepancy there was no explicit punishment for being caught. The reputation cost of being caught cheating was not made explicit. Moreover, there were no real or announced consequences for letters of references that were provided at the end of the contract.

3.3.2.2. Round 2:

Building on the experience from the first round, the second round of experimentation was conducted more systematically. First, the daily audit rate was explicitly randomized by day. The monitoring rates varied from zero to 100 percent. The frequency of monitoring rates is shown in Figure 3.2. Lower probability audit rates were assigned with greater frequency. The second round of data collection involves 21 distinct interviewers across a period of 17 days. As in the first round, not all interviewers work each day; the total number of observations is 306.

Specific details pertaining to the timeline of activities for the work day for this round of data collection are also presented in Figure 3.1. At the beginning of each work day interviewers received either 150 or 180 Malawian kwacha; a set number of condoms; and questionnaires. As in the first round of data collection, interviewers were informed as to how many in the group would be audited each day. Each day the randomly determined probability of audit is the same for all employees. However in this case the announcement was made by the experimenter rather than the supervisors. During the work day, at the end of each interview interviewers were required to give the respondent 30 Malawian Kwacha and then offer the respondent the opportunity to purchased reduced-priced condoms, just as in round one.

At the end of each work day interviewers were required to submit all questionnaires to their supervisor. All remaining condoms and coins were handed in to the experimenter the following morning. On days when there was no monitoring, all employees were asked to place all remaining stock of condoms and coins in one large sealed box. This was done publicly and given that each bag was unmarked it was clear that interviewer bags of coins and stock could not be traced to a specific interviewer. On days with 100-percent monitoring, each interviewer submitted their stock and coins. These were labeled and reconciled later in the day. On days with partial monitoring, first a random draw of interview IDs took place. Each interviewer ID was placed in a black bag. A number of ID tokens corresponding to the day's audit rate were drawn from the bag, in full view of all interviewers. For those interviewers who were selected for auditing, the same process as on days of full monitoring occurred. For those interviewers who were not selected the same protocol for days of no monitoring was followed.

After reconciling the stock, coins, and questionnaires, the experimenter communicated discrepancies to individual interviewers in private the following day. Individuals who had not cheated were informed that they had not cheated.

3.3.2.3 Key differences in data collection by round

Several differences in implementation between rounds one and two may affect interpretation of the results. First, the audit rate was explicitly randomly determined in round two. In round one, it was determined in an ad-hoc basis unrelated to the prevalence of crime, but it was not explicitly randomized. Second, supervisors were not an intermediary in the collection process of the stock and coins in round two. Therefore, in round two all responsibility lies with the interviewer and there is no possibility for collusion with the supervisors. Third, in round one while the daily audit rate was announced all interviewers were in fact audited. In round two, the audit rate that was announced was implemented. The random selection of interviewers selected for auditing was also conducted in an open transparent manner. Therefore, the round two process was likely more credible and transparent. Fourth, there is considerably less variation in the partial monitoring rates in round one as compared to round two.

While the implementation of the experiment differed on those dimensions, many design elements were preserved across the two rounds. These key similarities include the type of work and scope for theft; the private notification of discrepancies; and the lack of explicit punishment when cheating was detected.

3.3.3 Data

Measures of employee crime are drawn from the daily administrative data from two rounds of the experiment. I complement this administrative data with a limited set of demographic data. In the first round, there are 13 distinct interviewers (all men) and 19 different days with non-missing data. In the second round of data, there are 21 distinct interviewers (76.1 percent are men) and 17 days of data. On any particular day there may be missing data for a subset of interviewers. There are a number of reasons why interviewers did not work all days of the contract. First, not all participants worked the full duration of the project. In some cases, they experienced early termination of their contract by their project field supervisor for reasons unrelated to the monitoring experiment. Second, some employees terminated their contracts early in all cases immediately beginning on another short term contract. Third, for a subset of interviewers in the first round of data collection, they were only hired mid-project. Lastly, in both rounds of data collection interviewers were absent for work for a variety of reasons, in all cases informing their supervisors of their absence prior to the morning revelation of the monitoring rate. In total, I use 499 interviewer-day observations in the analysis, 193 observations from round one and 306 observations from round two.

Table 3.1 presents descriptive statistics for each round of data collection and the pooled dataset. As all employees in round one were male and 76.1 percent of the employees in round two were male, only 14.7 percent of employees in the pooled sample were female. All participants had completed their secondary schooling as this was an eligibility requirement for the job. Interviewers were on average 25 years of age; in round one 36 percent were married and in round two 25 percent of the employees were married. The predominant ethnicity in the area is the Yao tribe and this is reflected in the ethnicity of the interviewers. In round one, 36.5 percent of the interviewers are Yao and 44.7 percent are Yao in round two. Table 3.1 also presents the summary statistics from the administrative data of the day-to-day work undertaken by the interviewers. These statistics show key differences between the two rounds of data collected. First, interviewers completed on average four and two interviews per day in rounds one and two respectively. The second round of data collection was a follow-up study and required finding the same respondents as interviewed during a baseline survey. Therefore, interviewers needed to allocate time to locating individuals who had moved thus less time in any particular day was available for conducting interviews. In both rounds of data collection interviewers worked six days of the week – Monday through Saturday.

3.4 Empirical strategy

In this section, I first discuss the measurement of the key outcome variables: whether interviewers stole and how much they stole. Then I present the main empirical specification and discuss the validity of the underlying assumptions for this empirical strategy.

3.4.1. Measuring employee crime

I use two key outcomes to measure employee crime. First, I construct a binary indicator equal to one if the interviewer returned too little money (Any money missing) and zero otherwise. Table 3.1 shows that in 45.1 percent of cases in round one and 60.9 percent of cases in round two, interviewers pocketed or miscounted some of the petty cash in a manner that resulted in a net loss to the employer. It is interesting to highlight that in 13 percent of cases in round 1 and 16.1 percent of cases in round two, interviewers actually returned excess cash. For the purposes of this binary indicator I code these interviewers as zeros (i.e. no money missing). The existence of excess money returned highlights the fact that the measure of employee crime also embodies calculation errors. However, the prevalence of money missing interviewer-day level observations is substantially higher than the frequency of interviewer-days in which excess money is returned.

Second, I construct a continuous variable equal to the discrepancy between the amount of coins that the interviewer returned and the amount that they should have returned (Amount of money missing). This is coded as negative in the cases where excess money is returned. Table 3.1 shows that conditional on not returning exactly the correct amount of cash, the average amount stolen is 14 and 7.4 Malawian Kwacha per person in rounds one and two respectively. This average includes the negative amounts, i.e. the cases when excess money is returned. This is 4.7 and 6.5 percent of the total money issued to an interviewer each day – a non-trivial share of the money each interviewer handled daily. Figure 3.3 presents the average amount stolen on days with no monitoring, partial monitoring and full monitoring for each round of data and in the pooled data.

Despite the differences in implementation across the two rounds, the patterns of employee crime are quite consistent. Specifically, in round one, 50 percent of interviewers stole any money when there was no chance of an interviewer audit; 43 percent under a partial monitoring regime, and 32 percent when all interviewers were audited. In round two, rates of theft are somewhat higher than round one under both full and partial monitoring regimes, but somewhat lower under the full monitoring treatment. This may be attributable to the more credible and transparent auditing process conducted in round two.

Figure 3.4 shows the average amount stolen under the three monitoring regimes and two rounds of data collected. On average, MK 10 was stolen in either round one or round two when there was no auditing. With complete monitoring, the average amount stolen fell to MK 3 in each round. Under partial monitoring there are larger differences in the amount stolen between the two rounds. In round one, interviewers stole approximately 9 MKW and in round two 4 MKW. The larger range of partial monitoring in round two may be a key factor for these differences.

The full distribution of the amount stolen for each of the different monitoring rates, using data from both rounds, is shown in Figure 3.5. We see that when full monitoring was implemented, a larger fraction of interviewers return the correct amount of money to the project. We can reject the null hypothesis that the distributions are equivalent using the Kolmogorov-Smirnov two-sample test for equality of distributions (p-value is 0.000). Similarly, using the Kolmogorov-Smirnov two-sample test for equality of distributions for the amount stolen under no monitoring and partial monitoring the p-value is 0.000 and thus the null hypothesis that these distributions are the same is also rejected.

3.4.2 Main empirical specification

In this paper, I estimate the causal impact of the exogenously-assigned audit rate on the prevalence and extent of employee crime as measured by money stolen by employees. I estimate the following specification:

$$C_{id} = \alpha + \beta M_d + \gamma X_i + \delta Y_w + \varepsilon_{id}$$

where C_{id} measures employee crime of individual i on day d. I use two measures of crime: a binary indicator equal to one if the interviewer i on day d returned too little money (i.e. any money missing) and zero otherwise, and a continuous measure capturing the amount of money missing by interviewer i on day d. For the continuous measure, the amount of excess money returned is coded as a negative value. M_d is the announced probability of an audit. This probability is the same across interviewers within a day. Therefore, variation in the audit rate is only present across time. X_i denotes interviewer fixed effects and Y_w captures day of the week fixed effects.

The identifying assumption is that the error term is uncorrelated with the monitoring variable. This is satisfied by the exogenous assignment of the audit probability. One test of this assumption is to test whether the audit rates are predicted by interviewer characteristics. In round two, there are numerous audit rates implemented with relatively small sample sizes; as such, I restrict the comparison across treatment arms for the purposes of this test to: zero, partial and full monitoring. Table 3.2 presents the mean of each baseline characteristic of the interviewers for each of these monitoring regimes for each round and for the pooled data. In Columns 4, 8 and 12, I present the p-value of the joint F-test of the significance of the coefficients on full and partial monitoring from a regression of the baseline characteristic on indicators for the full and partial monitoring regimes. Table 3.2 shows that the different treatment arms appear to be balanced across the limited set of baseline characteristics available. In most cases, I cannot reject the null hypothesis that the means are jointly equal across the groups. One exception is the average age across groups in the pooled data. Interviewers in round two are on average slightly younger than those employed in round one by approximately one year; and in round two a larger fraction of the observations fall under partial monitoring. Therefore this systematic difference in the average age across monitoring regimes in the pooled data is a direct consequence of these factors. The marginally significant difference in the fraction male across monitoring regime in the pooled data arises due to there being no female interviewers in round one and higher frequency of partial monitoring in round two.

Given the random assignment and the validity of the underlying assumption that the error term is indeed uncorrelated with the monitoring rate, β_1 can be interpreted as a causal estimate of the impact of the monitoring rate on actual employee crime. The main results come from linear probability models for ease of interpretation, but probit models yield qualitatively similar results for the case where the binary indicator of employee crime is used. There are limited additional covariates available, including the age, tribe and marital status of the interviewer; whether or not these additional covariates are included in the regression does not substantively affect the results.

The usual assumption that ε_{id} is independent and identically distributed is violated due to the fact that we have multiple observations per person, and multiple observations per day. As different interviewers likely exhibit different cheating norms, I include interviewer fixed effects. Also, it is reasonable to assume that observations of one interviewer are correlated and therefore important to cluster the standard errors by interviewer. Since treatment varies by day, it is appropriate to cluster the standard errors at that level. To account for both clustering at the day and interviewer level simultaneously I implement two-way clustering (Cameron, Gelbach and Miller, 2008) in all regressions.

$$C_{id} = \alpha + \beta M_d + \gamma_i X_i + \delta_w Y_w + \varepsilon_{id}$$

where: C_{id} measures employee crime of individual i on day d. To measure the employee crime I use a binary indicator equal to one if the individual i on day d returned too little money (i.e. stole any money) and zero otherwise.I also use a continuous measure capturing the amount of money stolen by individual i on day d. M_d is the probability of an audit that employees are told. This probability is the same across individuals within a day. Therefore, variation is only induced across time not across individuals. X_i denotes individual fixed effects and Y_w captures day of the week fixed effects.

Given that M_d is randomly determined by the experimenter the audit rate is exogenously determined unlike most other studies. Therefore, it is assumed that the error term is uncorrelated with the monitoring variable. One test of this assumption is to compare observable characteristics across the different treatment arms. In round two, there are numerous audit rates implemented with relatively small sample sizes, as such I restrict the comparison across treatment arms for the purposes of this test to: no monitoring, partial monitoring and full monitoring. Table 3.2 presents the mean of each baseline characteristic of the interviewers for each of these monitoring regimes for each round and for the combined data. In Columns 4, 8 and 12, I present the p-value of the joint F-test of the significance of the coefficients on full monitoring and partial monitoring from a regression of the baseline characteristic on indicators for full and partial monitoring regimes. Table 3.2 shows that the different treatment arms appear to be balanced across the limited set of baseline characteristics available. In most cases, I cannot reject the null hypothesis that the means are jointly equal across the groups. In both rounds, there is no statistically different rate of monitoring observed across interviewer characteristics such as marital status, age and ethnicity. Only in one case when the data from the two rounds are combined there is a systematic difference by baseline characteristics across the monitoring regimes. Specifically, the average age of participants systematically differs across the monitoring regimes. Individuals in round two are on average slightly younger than those employed in round one by approximately one year; and in round two a larger fraction of the observations fall under partial monitoring. Therefore this systematic difference in ages in the combined data is a direct consequence of this combination of factors.

Given the random assignment and the validity of the underlying assumption that the error term is indeed uncorrelated with the monitoring rate, β_1 can be interpreted as a causal estimate of the impact of the monitoring rate on actual employee crime. OLS regression results are presented, however the probit results present qualitatively similar results for the case where the binary indicator of employee crime is used. There are limited additional covariates available, these include: the age, tribe and marital status of the interviewer; whether or not these additional covariates are included in the regression do not alter the results.

The usual assumption that ε_{id} is independent and identically distributed is violated due to the fact that we have multiple observations per person, and multiple observations per day. It is reasonable to assume that observations of one interviewer are correlated. An interviewer who steals money on one day is also likely to steal on other days. To control for this, all specifications will include interviewer fixed effects. Similarly, observations on one day are likely to be correlated as individuals faced similar conditions on each day. This is particularly likely given that the percentage monitored is determined for all interviewers each day rather than each interviewer being individually assigned a probability of being monitored. Therefore, in all regressions the standard errors will be clustered by day.

3.4.3 Results

Table 3.3 Panel A presents the impact of the monitoring rate on the rate of employee crime; and Panel B presents the impacts on the amount of money stolen by interviewers.

We observe that interviewers respond more strongly to monitoring in round two than in round one. In the first round, interviewers reduced stealing any money by between 16.7 and 21.7 percentage points when moving from a regime of no monitoring to full monitoring. The response is considerably higher in round two. Interviewers were approximately 55 percentage points less likely to have stolen money when everyone was audited compared to the case when no one was audited. Given that the second round of data collection was conducted in a more systematic manner, the second round estimates are my preferred estimates. However, in both cases the results are statistically and substantively significant and are robust to the inclusion of interviewer fixed effects and day of the week fixed effects.

Table 3.3 Panel B presents the impact of the monitoring rate on the amount of money stolen for round one (Columns 1 through 3), round two (Columns 4 through 6), and the pooled data from both rounds (Columns 7 through 9). Across the two rounds of implementation and across specifications, the results suggest a consistent response by interviewers. I find that increasing the monitoring rate from no monitoring to full monitoring reduced the amount of money stolen per person by five to nine kwacha. While in dollar terms this is a small amount of money, in percentage terms this is a large reduction, representing 50 to 80 percent of baseline theft depending on the round and empirical specification used. The estimated effect is not statistically significant at the 10 percent level in either round one or round two separately because of limited statistical power in the small samples. However, when pooling both rounds of data the reduction in the amount stolen is statistically significant at the 95 percent confidence level. Table 3.4 shows that there is considerable variation in the extent of corruption across interviewers. In round one, all interviewers have money missing on at least one day. The prevalence of money missing ranges from 10.5 percent to 88.8 percent of the time across interviewers. However, in round two 11 interviewers never return less money than they should. In this case, prevalence of money missing ranges from zero percent to 75 percent. The distribution of the amount stolen by interviewer also varies considerably.

Thus far, regression specifications have implicitly assumed that the relationship between crime and the monitoring rate is linear. Table 3.5 allows for a somewhat more flexible relationship between crime and the monitoring rate. Due to the limited variation in the partial monitoring rate in the first round of data used I examine two semi parametric specifications. Table 3.5 Column 1 includes a dummy indicator for the presence of any monitoring; while Column 2 includes a dummy for the presence of monitoring higher than 50 percent. These results are fairly similar in magnitude as any monitoring resulted in a large a reduction in crime. The fact that the magnitude of the coefficients in column 1 and 2 are so similar suggests also that the responsiveness is non-linear.

Round two has more variation in the partial monitoring rates and therefore permits a more systematic test of the linearity of the relationship between monitoring and crime. If the probability of crime exhibited a linear relationship with the monitoring rate then we would observe the magnitude of the coefficient increasing by a constant rate across the specifications in Columns 3 through 7. Table 3.5 Column 3 shows that any monitoring results in a 12.7 percentage point decline in crime; although large in magnitude the coefficient is not statistically significant at the 10 percent level. The coefficient becomes considerably larger at a monitoring rate of 20 percent (Column 3). In the presence of an audit rate of 20 percent or more, the decline in crime was approximately 35.9 percentage points. Although the magnitude of the coefficient does increase across specifications it does not do so in a linear manner. Therefore, it appears that there exists a non-linear relationship between the probability of employee crime and the monitoring rate.

Table 3.5 Panel B replicates Panel A for the continuous measure of corruption. Interestingly, I observe coefficients of roughly the same magnitude across specifications. This result suggests that any monitoring is sufficient to reduce the average amount of money stolen, and that there are no gains to increasing the monitoring rate beyond five percent in this setting.

3.5 Conclusion

In this paper I examine the impact of monitoring on theft by short-term contract employees in Malawi. I induce exogenous variation in the monitoring rate by explicitly varying the daily audit rate, enabling precise causal estimates of the impact of monitoring on employee crime. Two experimental rounds with slightly different protocols yield similar results.

I find that a ten percentage point increase in the probability of audit for an individual employee decreases the likelihood of money being stolen by four percentage points and reduces the amount stolen by between five and eight percent of baseline theft. I find that the relationship between the monitoring rate and probability of stealing is non-linear. This finding deserves further attention in future work as it can help determine optimal monitoring regimes.

Understanding the efficacy of monitoring on employee crime, opportunistic behavior and effort more generally is particularly relevant from a policy perspective. Developing countries are widely believed to have higher levels of aggregate corruption and indeed, economists believe the corruption is a key obstacle to aggregate growth (Shleifer and Vishny 1993, Mauro 1994). If employees respond to monitoring by reducing corruption, then monitoring may be part of a broader micro-approach enabling markets to function more effectively and countries to grow more rapidly. The results in this paper are a first attempt at shedding on light on the precise causal impact monitoring can have on reducing corruption. More work that explicitly examines the dynamics, and non-linearities of the relationship between monitoring and employee effort more broadly is needed.

3.6 References

Barnes, P. and Lambell, J. (2007). Organizational Susceptability to Fraud: Does Fraud Strike Randomly or are there Organizational Factors Affecting its Likelihood and Size?, *Managerial and Decision Economics*, Volume 28, Issue 3, pages 181–193.

Becker, G. and Stigler, G.J. (1974), "Law Enforcement, Malfeasance, and Compensation of Enforcers," *The Journal of Legal Studies*, Vol. 3, No. 1 pp. 1 – 18.

Chandra, V. et al. (2001), "Constraints to Growth and Employment in South Africa. Report No. 2: Evidence from the Small, Medium and Micro-Enterprise Firm Survey," *The World Bank Southern African Department Informal Discussion Papers on Aspects of the South African Economy Series.*

Chang, J.J. and Lai, C.C. (1999), "Carrots or sticks? A Social Custom Viewpoint on Worker Effort," *European Journal of Political Economy*, Vol. 15, pp. 297 – 310.

Dickens, William T, et al, (1989), "Employee Crime and the Monitoring Puzzle," *Journal of Labor Economics*, University of Chicago Press, Vol. 7(3), pp. 331-47, July.

di Tella, R. and Schargrodsky, E. (2003), "The Role of Wages and Auditing During a Crackdown on Corruption in the City of Buenos Aires," *Journal of Law and Economics*, Vol. 46, pp. 269 – 292.

Fafchamps, M. (2004), "Market institutions in Sub-Saharan Africa: Theory and evidence", Cambridge, Massachusetts: MIT Press.

Frey, B. S. (1993a),"Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty," *Economic Inquiry*, Oxford University Press, vol. 31(4), pages 663-70, October.

–––(1993b), "Shirking or Work Morale? The Impact of Regulating," *European Economic Review*, Vol. 37, pp. 1523 – 1532.

Groshen, E.L. and Krueger, A.B. (1990), "The Structure of Supervision and Pay in Hospitals," *Industrial and Labor Relations Review,* Vol. 43, pp. 134 – 146S.

Kruse, D. (1992), "Supervision, Working Conditions, and the Employer Size-Wage Effect," *Industrial Relations,* Vol. 31, pp. 229 – 249.

Lazear, E. (1996), "Performance Pay and Productivity," *NBER Working Paper* No. 5672.

Mauro, P. (1995), "Corruption and Growth," *Quarterly Journal of Economics*, Vol. 110(3), pp. 681 – 712.

Nagin, D.S. and Paternoster, R. (1993), "Enduring individual differences and rational choice theories of crime," *Law and Society Review*, Vol. 27, pp. 467–96.

Nagin, D.S., Rebitzer, J.B., Sanders, S. and Taylor, L.J. (2002), "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment." *American Economic Review,* Vol. 92, pp. 850–73.

National Bank of Malawi. (2009), "NBM Visa Card" [Online] Available at: http://www.natbank.co.mw/index.php?pagename=debit_card

Paarsch, H. and Shearer, B. (1996), "The Response of Worker Effort to piece rates: Evidence from the British Columbia Tree-Planting Industry," *The Journal of Human Resources*, Vol. 34, No. 4, pp. 643 – 667.

Prendergast, C. (1999), "The Provision of Incentives in Firms." *Journal of Economic Literature*, Vol. 37 No.1. pp. 7 – 63.

PriceWaterhouseCooper. (2002), "Report to the Nation on Occupational Fraud and Abuse".

Rebitzer, J.B. (1995), "Is There a Trade-Off Between Supervision and Wages? An Empirical Test of Efficiency Wage Theory," *Journal of Economic Behavior and Organization*, Vol. 28(1), pp. 107 – 129.

Rickman, N. and Witt, R. (2007), "The Determinants of Employee Crime in the UK." *Economica*. Vol. 74 pp 161 – 175.

Snyder, N.H. et al. (1991), "Reducing Employee Theft – A Guide to Financial and Organizational Controls," *Quorom Books* (New York).

Shleifer, A. and Vishny, R. (1993), "Corruption," *Quarterly Journal of Economics*, Vol. 108(3), pp. 599 – 617.

United Nations and South African Department of Public Service and Administration. (2003), "Country Corruption Assessment Report – South Africa," April 2003.

Figure 3.1: Timeline of rounds one and two of audit study

	Round 1	Round 2
t = -1	Experimenter counts and sorts coins into 300 MKW bags	Experimenter counts and sorts coins into 150 or 180 MKW bags
t = 0	Supervisors provide interviewers with coins, questionnaires and condoms. Supervisors announce the number of interviewers that will be audited that day. Interviewers work during the day completing interviews. At end of work day: interviewers hand in questionnaires to one supervisor, coins and condoms to other supervisor.	Supervisors provide interviewers with coins, questionnaires and condoms. Experimenter meets with employees, announces number of interviewers to be audited that day. Interviewers work during the day completing interviews. At end of work day: interviewers hand in questionnaires to one supervisor.
t = 1	Supervisors submit coins, and remaining condom stock to experimenter. Experimenter conducts audit - counts coins and compares numbers to completed questionnaires, and condoms sold. Experimenter leaves list of discrepancies for supervisors.	Experimenter randomly draws interviewers to be audited. These interviewers submit coins, and condom stock to experimenter. The remaining interviewers submit any coins and stock to the sealed box. Experimenter conducts audit and communicates discrepancies directly to the interviewer.
t = 2	Supervisors communicate (if monitoring rate >0) privately to interviewers their discrepancy.	

Notes: Each interviewer-day level observation is each day worked (t=0). However actions related to the implementation of the experiment for each day of data collected starts the day prior to the work day, and in the case of round one finishes two days after the work day, and one day after the work day for round two.

Figure 3.2: Audit rate assignment: round two

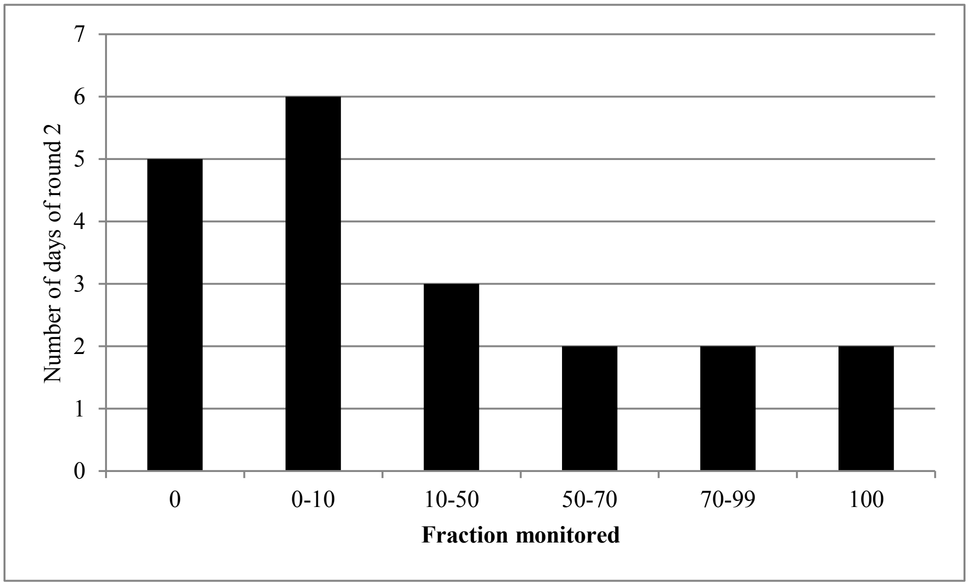


Figure 3.3: Monitoring Rates and Propensity to Steal

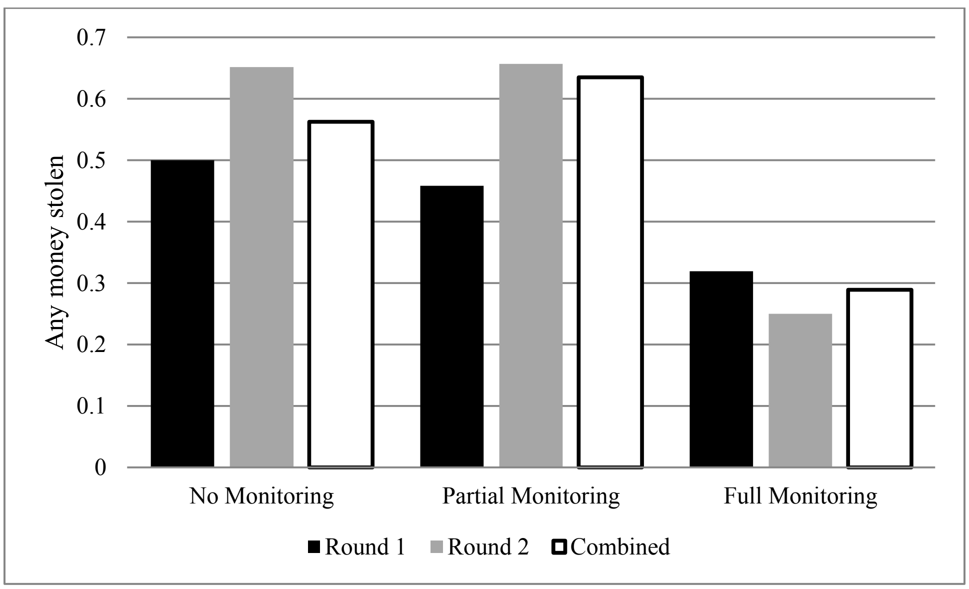


Figure 3.4: Monitoring Rates and Average Amount Stolen

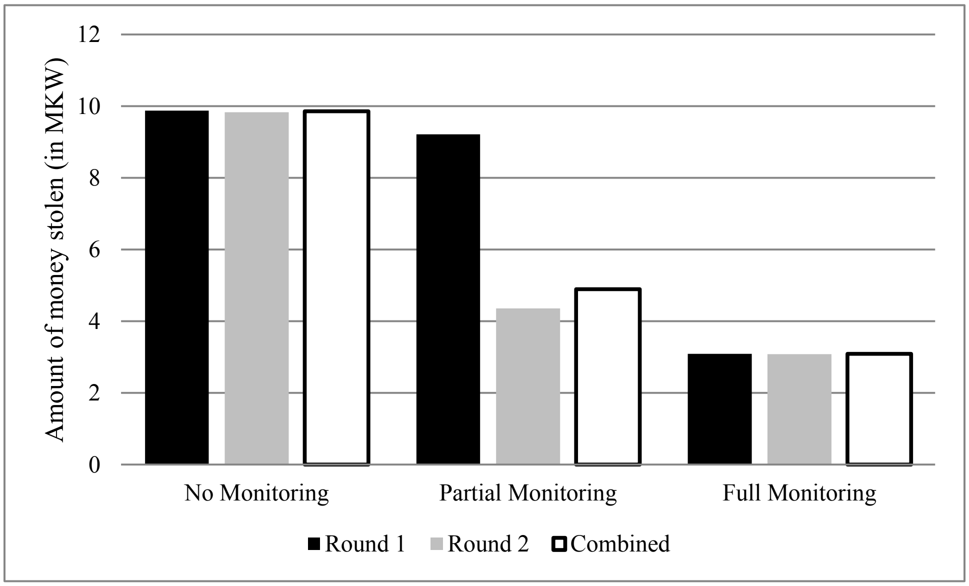


Figure 3.5: Cumulative distribution functions of amount of money stolen by monitoring regime

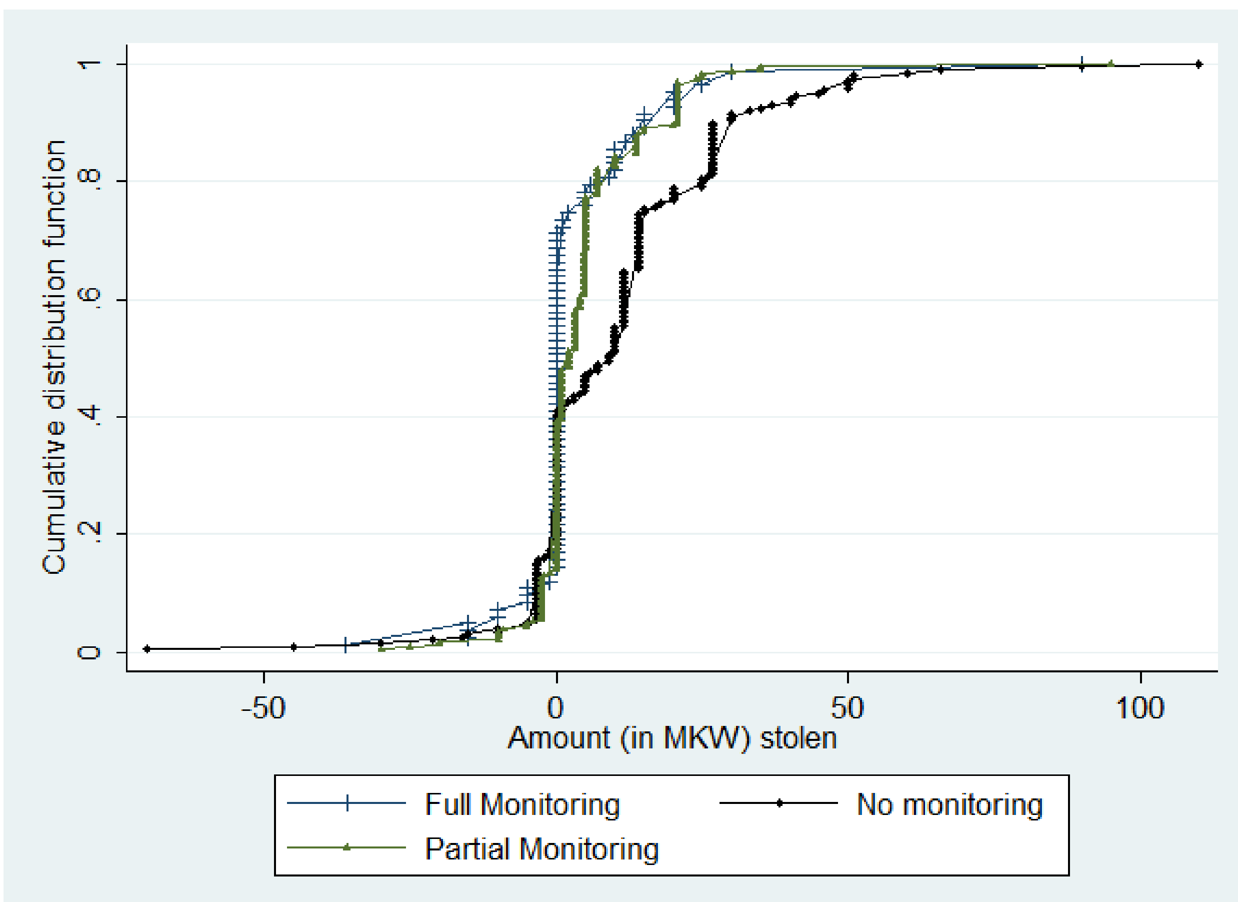


Table 3.1: Descriptive statistics

	Round 1 N=193		Round 2 N=289		Pooled N=499	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Interviewer characteristics:						
Male	1.000	0.000	0.773	0.419	86.055	0.347
Age	25.709	2.687	23.742	3.019	24.523	3.045
Married	0.358	0.481	0.254	0.436	0.296	0.457
MCSE (Completed secondary schooling)	1.000	0.000	1.000	0.000	1.000	0.000
Yao	0.365	0.483	0.447	0.498	0.416	0.493
Audit study administrative data:						
Percent monitored	0.285	0.421	0.315	0.363	0.303	0.386
Number of interviews completed	3.916	1.537	2.088	1.369	2.769	1.684
Number of daily monetary transactions	5.442	2.444	3.739	1.508	4.470	2.136
Value of coins distributed	298.653	5.026	155.794	31.376	209.436	73.611
Days of the week:						
Monday	0.093	0.292	0.237	0.426	0.183	0.387
Tuesday	0.202	0.403	0.178	0.383	0.187	0.390
Wednesday	0.218	0.414	0.178	0.383	0.193	0.395
Thursday	0.212	0.410	0.112	0.316	0.150	0.357
Friday	0.176	0.382	0.121	0.327	0.142	0.349
Saturday	0.098	0.299	0.174	0.380	0.174	0.380
Worker Effort:						
Minutes worked	304.370	119.677	105.300	203.653	285.321	142.056
Proportion of work day worked	0.644	0.228	0.355	0.437	0.616	0.268
Dependent variables:						
Any money missing	0.451	0.499	0.609	0.489	0.549	0.498
Amount stolen	8.140	21.016	5.698	9.532	6.622	14.980
Amount stolen (conditional on any money missing)	14.027	26.089	7.402	10.271	9.486	17.161
Returned excess money	0.130	0.337	0.161	0.368	0.149	0.356

Notes:

The data comes from two rounds of data collected. The round one data consists of 13 interviewers across 19 days; the second round consists of 21 interviewers across 17 days.

Table 3.2: Balanced on observables?

	Round 1				Round 2				Combined			
	No Monitoring	Partial Monitoring	Full Monitoring	p-value of joint F-test	No Monitoring	Partial Monitoring	Full Monitoring	p-value of joint F-test	No Monitoring	Partial Monitoring	Full Monitoring	p-value of joint F-test
Male	1.000	1.000	1.000	n/a	0.724	0.801	0.730	0.336	0.894	0.823	0.881	0.101
Age	25.703	25.583	25.787	0.953	23.652	23.805	23.576	0.882	24.947	24.010	24.875	0.005
Married	0.361	0.333	0.362	0.965	0.246	0.259	0.242	0.964	0.319	0.268	0.313	0.495
MCSE	1.000	1.000	1.000	n/a	1.000	1.000	1.000	n/a	1.000	1.000	1.000	n/a
Yao	0.373	0.333	0.362	0.933	0.434	0.459	0.405	0.807	0.397	0.445	0.381	0.474

Notes:

The p-value of the joint F-test comes from the test that the coefficinet on full and partial monitoring indicators equal zero, when the baseline characteristic is regressed on partial and full monitoring dummy indicator variables. MCSE refers to completed secondary schooling.

Table 3.3: Response to monitoring

Panel A: Any money missing

	Round 1			**Round 2**			**Combined**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Percent Monitoring	-0.180**	-0.167*	-0.217**	-0.555***	-0.560***	-0.530**	-0.383***	-0.382***	-0.441***
	[0.085]	[0.087]	[0.110]	[0.177]	[0.184]	[0.220]	[0.109]	[0.112]	[0.109]
Constant	0.502***	0.196*	0.215*	0.803***	0.692**	0.740***	0.371***	-0.129	-0.093
	[0.078]	[0.103]	[0.123]	[0.131]	[0.314]	[0.228]	[0.114]	[0.296]	[0.260]
Interviewer Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Day of the week Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
Observations	193	193	193	307	306	306	500	499	499
R-squared	0.023	0.167	0.212	0.172	0.211	0.293	0.119	0.195	0.234

Panel A: Amount of money missing

	Round 1			**Round 2**			**Combined**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Percent Monitoring	-6.741*	-4.866	-5.082	-6.949	-7.150	-9.462	-6.854***	-6.112***	-7.825**
	[3.595]	[3.762]	[5.183]	[4.280]	[4.392]	[6.452]	[2.253]	[2.301]	[3.089]
Constant	10.058***	3.769	3.110	8.241***	1.081	0.206	11.969***	7.444	5.484
	[2.340]	[4.688]	[6.375]	[3.069]	[2.359]	[3.858]	[4.016]	[7.122]	[8.001]
Interviewer Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Day of the week Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
Observations	193	193	193	307	306	306	500	499	499
R-squared	0.018	0.122	0.133	0.074	0.109	0.189	0.036	0.123	0.142

Notes:

Results presented are from the OLS specification. In the case of Panel A, using a probit produces qualitatively similar results to those presented here. Standard errors are clustered by both project day and interviewer. *** indicates significance at the 1 percent level, ** indicates significance at the 5 percent level, and * indicates significance at the 10 percent level. Any amount missing records a 0 if the employee returned too much or precisely the correct amount, and a 1 if the employee returned too little money. Amount stolen is not truncated it simply measures the discrepancy (positive, zero or negative) between what the employee should have and did return - there are cases in which the amount stolen is a negative number (i.e. the individual returned too much). Additional controls include: an ethnicity dummy for Yao tribe, married, and age.

Table 3.4: Interviewers' distributions

Panel A: Round 1

	Obs	Corrupt		Amount Stolen				Percent of rejections of Kolmogorov-Smirnov test of equality with other interviewers' distributions		
		Mean	SD	Mean	SD	Min	Max	1%	5%	10%
1	15	0.133	0.352	1.933	14.478	-21	50	0.083	0.250	0.417
2	17	0.588	0.507	9.000	15.447	-16	51	0.000	0.167	0.250
3	18	0.611	0.502	3.667	12.565	-30	20	0.083	0.167	0.167
4	18	0.555	0.511	6.055	10.602	-15	25	0.000	0.083	0.167
5	18	0.555	0.511	9.944	19.425	-15	66	0.000	0.083	0.250
6	13	0.461	0.519	2.923	21.150	-45	50	0.000	0.083	0.083
7	19	0.368	0.496	12.052	34.850	-70	110	0.000	0.083	0.083
8	18	0.500	0.514	2.222	13.653	-36	30	0.000	0.083	0.167
9	19	0.105	0.315	2.947	14.397	-10	60	0.167	0.417	0.500
10	13	0.461	0.519	11.923	16.520	0	45	0.000	0.000	0.000
11	9	0.888	0.333	29.550	36.613	0	95	0.167	0.500	0.667
12	4	0.500	0.577	32.500	42.720	0	90	0.000	0.000	0.000
13	12	0.333	0.492	10.080	18.630	0	50	0.000	0.083	0.083

Panel B: Round 2

	Obs	Corrupt		Amount Stolen				%Days Indiv Checked
		Mean	SD	Mean	SD	Min	Max	
1	5	0.400	0.548	2.800	4.087	0	9	0.294
2	5	0.000	0.000	-2.000	4.472	-10	0	0.313
3	4	0.500	0.577	7.500	11.902	0	25	0.235
4	6	0.000	0.000	-3.500	8.093	-20	0	0.353
5	8	0.375	0.518	3.625	8.863	-2	25	0.471
6	2	0.000	0.000	0.000	0.000	0	0	0.400
7	4	0.750	0.500	15.000	12.910	0	30	0.222
8	5	0.200	0.447	3.000	6.708	0	15	0.294
9	5	0.200	0.447	4.000	8.944	0	20	0.294
10	5	0.000	0.000	0.000	0.000	0	0	0.294
11	4	0.000	0.000	-1.500	2.380	-5	0	0.235
12	5	0.000	0.000	0.000	0.000	0	0	0.313
13	5	0.000	0.000	-0.800	1.789	-4	0	0.313
14	4	0.500	0.577	2.250	16.338	-20	15	0.235
15	5	0.400	0.548	5.000	7.071	0	15	0.294
16	7	0.286	0.488	-0.429	15.957	-30	25	0.412
17	7	0.143	0.378	0.714	17.661	-25	35	0.412
18	2	0.000	0.000	0.000	0.000	0	0	0.333
19	6	0.000	0.000	0.000	0.000	0	0	0.353
20	6	0.000	0.000	0.000	0.000	0	0	0.353
21	3	0.000	0.000	0.000	0.000	0	0	0.429

Table 3.5: Non-linearities in the response to monitoring

Panel A: Any Money Missing

	Round 1		**Round 2**					**Combined**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
5% chance or greater	-0.181*		-0.127					-0.179				
	[0.108]		[0.223]					[0.125]				
10% chance or greater				-0.199					-0.244**			
				[0.176]					[0.120]			
20% chance or greater					-0.359**					-0.353***		
					[0.173]					[0.101]		
33% chance or greater						-0.396**					-0.368***	
						[0.170]					[0.100]	
50% chance or greater		-0.198**					-0.396**					-0.381***
		[0.099]					[0.170]					[0.091]
Constant	0.201	0.206*	0.507	0.455	0.561**	0.523**	0.523**	-0.117	-0.053	-0.003	0.089	-0.047
	[0.126]	[0.119]	[0.443]	[0.298]	[0.252]	[0.244]	[0.244]	[0.230]	[0.218]	[0.226]	[0.253]	[0.245]
Interviewer Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of the week Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	193	193	316	316	316	316	316	509	509	509	509	509
R-squared	0.209	0.209	0.208	0.231	0.280	0.300	0.300	0.160	0.188	0.234	0.243	0.242

Panel B: Amount of money missing

	Round 1		Round 2					Combined				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
5% chance or greater	-3.796		-7.390					-6.189*				
	[4.848]		[6.005]					[3.514]				
10% chance or greater				-6.995*					-6.044**			
				[4.017]					[2.578]			
20% chance or greater					-7.936					-6.182**		
					[4.914]					[2.626]		
33% chance or greater						-6.731					-5.826***	
						[4.096]					[2.158]	
50% chance or greater							-6.731					-6.416***
							[4.096]					[2.325]
Constant	2.575	3.007	4.679	-0.227	1.032	-0.572	-0.572	1.767	3.087	3.282	4.497	2.436
	[6.299]	[6.230]	[7.393]	[3.308]	[4.209]	[3.384]	[3.384]	[8.035]	[8.710]	[8.769]	[8.713]	[8.648]
Interviewer Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of the week Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	193	193	316	316	316	316	316	509	509	509	509	509
R-squared	0.130	0.133	0.177	0.190	0.185	0.158	0.158	0.139	0.142	0.141	0.137	0.141

Notes:

Results presented are from the OLS specification. In the case of Panel A, using a probit produces qualitatively similar results to those presented here. Standard errors are clustered by both project day and interviewer. *** indicates significance at the 1 percent level, ** indicates significance at the 5 percent level, and * indicates significance at the 10 percent level. Any amount missing records a 0 if the employee returned too much or precisely the correct amount, and a 1 if the employee returned too little money. Amount stolen is not truncated it simply measures the discrepancy (positive, zero or negative) between what the employee should have and did return - there are cases in which the amount stolen is a negative number (i.e. the individual returned too much). Additional controls include: an ethnicity dummy for Yao tribe, married, and age.

APPENDIX A

Determinants of the hiring decision using administrative data

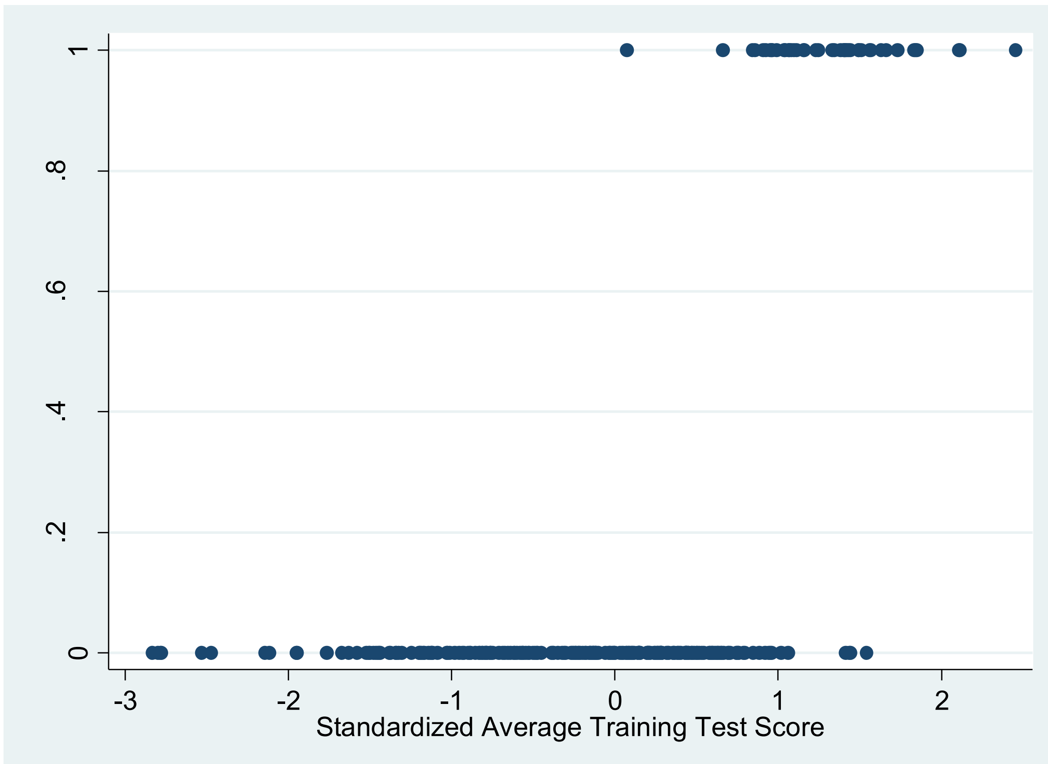
In making hiring decisions the recruiter took a number of factors into account. As discussed above the recruiter conducted multiple tests to ensure that trainee participants paid attention and to ensure an objective measure of assessment was available to them. No participants were hired that had a standardized test score (using the composite test measure) less than 0.05. All participants that had a standardized test score greater than 1.3 were hired. As such although performing well on the test is a key factor in the hiring decision, 72 percent of the group that were hired had test scores in a region where that was not a sufficient determining factor. That is, performing well on the test was a necessary condition to get hired. It was not however a sufficient condition for those participants with a standardized test score between 0.05 and 1.3.

Appendix Table A.1 presents the determinants of the hiring decision making process of the recruiter. This shows that the standardized test score is an important determinant of whether the person gets hired - a 1 standard deviation increase in the composite test score results in 9.7 percentage point increase in the likelihood that the individual is hired. Other key indicators that were measured by the recruiter include punctuality, contributions and disruptions. Given Appendix Figure A.1, it suggests that any alternative measures of evaluating performance should be interacted with the test score.

Punctuality appears to have little impact on the hiring decision. Interestingly for those individuals that do come late, this seems to increase their chances of employment if they have higher standardized tests scores (although the magnitude is small – for every additional minute late they are 0.3 percentage points more likely to be hired if they have a standardized test score of 1) (Column 2 of Appendix Table 2). Appendix Table 2 also shows that for those performing well (in terms of their standardized test score), making “good” and “neutral” contributions during the training sessions increased the probability that they were hired. In such a large hiring process being noticed in a good way mattered for those participants who performed well but not exceptionally well. Lastly, Column 4 of Appendix Table 2 also includes measures for disruptions made by participants during the training. This appears not to have any significant impact on the hiring decision making process as the magnitude of the coefficients are small and statistically insignificant.

Evidently, the most significant factor taken into account by the recruiter in its hiring decisions was the performance of participants on the written tests. However, there is evidence that other performance indicators were also taken into account – in particular whether or not the applicant made a “good” contribution to the discussion.

Figure A.1: Scatter plot employed by recruiter and training test score



Appendix Table A.1: Predicting Employment

	(1)	(2)	(3)	(4)	(5)
Age	0.107***	0.097***	0.069***	0.065***	-0.007
	[0.018]	[0.016]	[0.016]	[0.016]	[0.005]
Married	0.036	-0.008	-0.007	-0.007	0.101
	[0.071]	[0.005]	[0.005]	[0.005]	[0.070]
Ever worked	0.067	0.086	0.104	0.103	0.093
	[0.058]	[0.069]	[0.067]	[0.069]	[0.059]
Ever worked with recruiter	0.150	0.096*	0.087	0.093	0.117
	[0.094]	[0.055]	[0.058]	[0.059]	[0.078]
Ability score (standardized)	0.104***	0.139*	0.122	0.12	0.046**
	[0.024]	[0.082]	[0.078]	[0.078]	[0.023]
Test score		0.097***	0.092***	0.067***	0.063***
		[0.016]	[0.016]	[0.016]	[0.016]
Minutes late			-0.035	-0.035	0.001
			[0.043]	[0.043]	[0.001]
Minutes late X test score			0.114**	0.096*	0.001
			[0.052]	[0.051]	[0.002]
Any good contribution				-0.031	-0.031
				[0.043]	[0.043]
Any good contribution X test score				0.114**	0.098*
				[0.052]	[0.052]
Any neutral contribution				0.023	0.023
				[0.042]	[0.042]
Any neutral contribution X test score				0.078	0.068
				[0.052]	[0.050]
Any bad contribution				-0.012	0.019
				[0.041]	[0.055]
Any bad contribution X test score				0.062	-0.052
				[0.041]	[0.061]
Any disruption					-0.009
					[0.041]
Any disruption X test score					0.059
					[0.042]
Constant	0.281**	0.272**	0.269**	0.250*	0.240*
	[0.137]	[0.129]	[0.130]	[0.137]	[0.145]
Observations	268	268	268	268	268
R-squared	0.11	0.25	0.26	0.31	0.32
Average of dep variable			0.158		

Notes: The dependent variable is a binary indicator equal to 1 if the recruiter offered the job-seeker a job and 0 otherwise. For covariates with missing data the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors. *** indicates significance at the 1% level, ** indicates significance at the 5% level, * indicates significance at the 10% level

APPENDIX B

Robustness checks I

This appendix presents additional specification checks for the analysis conducted in Chapter 1: Employee Risk and Performance. Appendix Table B.1 considers the additional outcome of training behavior. Appendix Table B.2 examines the distributions of arrival times at the training venues for individuals assigned to different treatment groups. Appendix Table B.3 and B.4 present the main effort and performance results when covariates are excluded. Appendix Table B.5 through Table B.9 present various robustness specification checks.

Appendix Table B.1: Training behavior by treatment group

Dependent Variable	Any disruption	# disruptions	Chat/ Noise	Toilet/ Move	Phone Call
	(1)	(2)	(3)	(4)	(5)
0% Job Guarantee	0.627	1.104	0.533	0.504	0.067
	[0.105]	[0.242]	[0.134]	[0.148]	[0.052]
1% Job Guarantee	0.696	1.120	0.536	0.488	0.096
	[0.104]	[0.203]	[0.099]	[0.135]	[0.047]
5% Job Guarantee	0.615	0.933	0.503	0.330	0.100
	[0.110]	[0.191]	[0.150]	[0.102]	[0.044]
50% Job Guarantee	0.586	0.887	0.285	0.419	0.183
	[0.105]	[0.190]	[0.082]	[0.121]	[0.072]
75% Job Guarantee	0.579	0.816	0.437	0.264	0.114
	[0.137]	[0.231]	[0.139]	[0.135]	[0.058]
100% Job Guarantee	0.638	1.021	0.560	0.468	-0.007
	[0.159]	[0.278]	[0.194]	[0.203]	[0.014]
Observations	268	268	268	268	268
R-squared	0.432	0.351	0.282	0.225	0.123
Stratification cell FE's?	Yes	Yes	Yes	Yes	Yes
Includes controls?	Yes	Yes	Yes	Yes	Yes
p-values of F-tests:					
0% and 100%	0.952	0.821	0.907	0.886	0.168

Notes:

This table presents the average training classroom behavior by treatment group using administrative data. "Any disruption" is a binary indicator equal to 1 if the job trainee at any point during training disrupted the training to exit the room, to take a phone call or was disruptive by talking to his peers or making noise. "Number of disruptions" is the cumulative number of disruptions made by a job trainee. Treatment status was randomly allocated and stratified by quintile ability and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Appendix Table B.2: Arrival time distribution tests of equality

Minutes early/late: p-value of kolmogorov smirnov distrinbution test of equality

	0% Job Guarantee	1% Job Guarantee	5% Job Guarantee	50% Job Guarantee	75% Job Guarantee	100% Job Guarantee
0% Job Guarantee		0.178	0.272	0.436	0.616	0.408
1% Job Guarantee			0.995	0.421	0.196	0.38
5% Job Guarantee				0.572	0.475	0.269
50% Job Guarantee					0.769	0.193
75% Job Guarantee						0.13
100% Job Guarantee						

Notes:

Arrival times were recorded by recruitment staff as discussed in Section 4.2. This table presents the associated p-values from Kolmogorov Smirnov distribution tests of equality between the distribution of arrival times between treatment groups.

Appendix Table B.3: Performance Indicators (No covariates)

	Tests	Contributions				
		Any	**# total**	**# good**	**# neutral**	**# bad**
	(1)	(2)	(3)	(4)	(5)	(6)
0% Job Guarantee	-0.176	0.635	1.453	0.491	0.604	0.358
	[0.147]	[0.068]	[0.212]	[0.088]	[0.122]	[0.108]
1% Job Guarantee	-0.015	0.611	1.589	0.804	0.589	0.196
	[0.136]	[0.067]	[0.256]	[0.134]	[0.127]	[0.086]
5% Job Guarantee	0.041	0.725	1.596	0.692	0.692	0.212
	[0.132]	[0.063]	[0.219]	[0.152]	[0.128]	[0.057]
50% Job Guarantee	0.041	0.642	1.389	0.778	0.389	0.222
	[0.124]	[0.067]	[0.219]	[0.139]	[0.093]	[0.063]
75% Job Guarantee	-0.039	0.741	1.321	0.750	0.500	0.071
	[0.241]	[0.085]	[0.234]	[0.150]	[0.120]	[0.049]
100% Job Guarantee	0.259	0.760	2.240	0.920	1.040	0.280
	[0.195]	[0.086]	[0.414]	[0.198]	[0.239]	[0.107]
Observations	258	262	268	268	268	268
R-squared	0.013	0.676	0.475	0.380	0.342	0.161
p-value of F-test:						
0% and 100%	0.076	0.254	0.092	0.048	0.105	0.607

Notes:

This table presents mean performance using an average across training for each job trainee. I use the average of the standardized test scores which are standardized by using the sample mean and standard deviation for the relevant test. "Any contribution" is a binary indicator if the job trainee engaged verbally ever during training. The "total number of contributions" is the cumulative number of contributions made by the job trainee during the whole training, and then separated out by quality as determined by the recruitment staff. Treatment status was randomly allocated and stratified by quintile ability and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Appendix Table B.4: Average Effort Indicators (No Covariates)

	Ever late	Always late	Mins early or late	Studied (Hours)	Radio/TV (Hours)
	(1)	(2)	(3)	(4)	(5)
0% Job Guarantee	0.180	0.020	-24.230	1.177	1.142
	[0.055]	[0.020]	[2.228]	[0.137]	[0.121]
1% Job Guarantee	0.182	0.000	-21.467	1.151	1.580
	[0.053]	[0.000]	[1.794]	[0.109]	[0.132]
5% Job Guarantee	0.314	0.020	-19.209	0.959	1.358
	[0.066]	[0.020]	[2.310]	[0.100]	[0.154]
50% Job Guarantee	0.176	0.020	-21.843	1.093	1.520
	[0.054]	[0.020]	[2.099]	[0.098]	[0.138]
75% Job Guarantee	0.259	0.037	-19.914	1.125	1.419
	[0.085]	[0.037]	[3.023]	[0.134]	[0.162]
100% Job Guarantee	0.280	0.080	-19.320	0.754	2.020
	[0.091]	[0.055]	[4.354]	[0.078]	[0.247]
Observations	259	259	259	254	254
R-squared	0.238	0.043	0.647	0.699	0.676
p-value of F-test:					
0% and 100%	0.347	0.306	0.316	0.008	0.002

Notes:

This table presents the average effort by treatment group using both administrative data and survey data. "Ever late" is a binary indicator equal to 1 if the job trainee ever arrived late for training. "Always late" is a binary indicator if the job trainee arrived late for training every day. "Minutes early/late" is a continuous variable recording the average minutes early (-) or late (+) job trainees arrived across the training period. Time use in columns 4 and 5 comes from survey data and is the average number of hours conducting each activity. Treatment status was randomly allocated and stratified by quintile ability and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Appendix Table B.5: Omitted variable bias ratio

	Ratio (1)
Performance indicators:	
Tests	67.994
Engagement:	
* # of contributions	-1.504
* # good contributions	-1.672
* # neutral contributions	-1.796
Effort indicators:	
Time use:	
* Hours studied training materials	7.003
* Hours watching tv/listening to radio	9.668

Notes:
Following Altonji et al. (2005) and Bellows and Miguel (2008), I construct a ratio that assesses the extent of omitted variable bias that would be required to explain away the results. This table presents the ratios for each of the performance and effort indicators for the estimated difference between those assigned no outside option and a guaranteed outside option. The ratio measures the extent to which selection on unobservables would need to exceed selection on observables to explain away the coefficient. Therefore, a larger ratio implies that the relative omitted variable bias from unobservables relative to observables is greater, and therefore estimated effects are less likely to be explained away.

Appendix Table B.6: Average performance by treatment group: Weighted results and bounds

	Average test score			Number of contributions			Good quality contributions		
		Min-Max Bounds			Min-Max Bounds			Min-Max Bounds	
	Weighted	0-75=max; 100=min	0-75=min; 100=max	Weighted	0-75=max; 100=min	0-75=min; 100=max	Weighted	0-75=max; 100=min	0-75=min; 100=max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0% Job Guarantee	-0.174	-0.067	-0.288*	0.629	0.652	0.618	0.242	0.259	0.24
	[0.141]	[0.147]	[0.154]	[0.104]	[0.110]	[0.104]	[0.051]	[0.056]	[0.051]
1% Job Guarantee	-0.004	0.045	-0.1	0.683	0.782	0.66	0.371	0.414	0.359
	[0.126]	[0.129]	[0.139]	[0.111]	[0.129]	[0.109]	[0.061]	[0.067]	[0.060]
5% Job Guarantee	0.038	0.089	-0.052	0.72	0.78	0.701	0.342	0.393	0.332
	[0.120]	[0.120]	[0.139]	[0.101]	[0.107]	[0.100]	[0.069]	[0.079]	[0.068]
50% Job Guarantee	0.03	0.16	-0.049	0.585	0.63	0.573	0.351	0.38	0.346
	[0.122]	[0.133]	[0.125]	[0.090]	[0.098]	[0.088]	[0.063]	[0.067]	[0.062]
75% Job Guarantee	-0.032	0.04	-0.156	0.503	0.555	0.485	0.297	0.341	0.287
	[0.208]	[0.218]	[0.232]	[0.097]	[0.105]	[0.094]	[0.067]	[0.077]	[0.065]
100% Job Guarantee	0.261	0.252	0.271	0.915	0.901	0.918	0.391	0.382	0.393
	[0.198]	[0.202]	[0.194]	[0.167]	[0.168]	[0.167]	[0.089]	[0.091]	[0.088]
Observations	258	268	268	262	268	268	262	268	268
R-squared	0.2	0.18	0.18	0.49	0.48	0.48	0.42	0.41	0.41
Stratification cell FEs?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p-value of F-test:									
0% and 100%	0.075	0.203	0.024	0.151	0.219	0.131	0.148	0.249	0.133

Notes:

This table presents mean performance using an average across training for each job trainee. I use the average of the standardized test scores which are standardized by using the sample mean and standard deviation for the relevant test. "Any contribution" is a binary indicator if the job trainee engaged verbally ever during training. The "total number of contributions" is the cumulative number of contributions made by the job trainee during the whole training, and then separated out by quality as determined by the recruitment staff. Treatment status was randomly allocated and stratified by quintile ability and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Appendix Table B.7: Lee bounds

	Lower Bound		Upper Bound		Trimming
	Coeff	p-value	Coeff	p-value	Proportion
	(1)	(2)	(3)	(4)	(5)
Performance indicators:					
Tests	0.346	0.154	0.492	0.054	5.66
Engagement:					
* Any contribution	0.1207	0.273	0.14	0.208	1.89
* Total # contributions	0.813	0.173	0.986	0.093	1.89
* # good contributions	0.413	0.192	0.509	0.099	1.89
* # neutral contributions	0.497	0.143	0.593	0.075	1.89
* # bad contributions	-0.155	0.429	-0.116	0.547	1.89
Effort indicators:					
Punctuality:					
* Always late	0.005	0.945	0.065	0.288	5.66
* Ever late	0.057	0.615	0.117	0.290	5.66
* Minutes early/late	1.894	0.709	6.490	0.206	5.66
Time use:					
* Hrs studied training materials	-0.502	0.001	-0.363	0.021	9.43
* Hrs watching tv/listening to radio	0.656	0.032	1.043	0.001	9.43

Notes:

This table presents the Lee bounds for the comparison of those assigned no outside option (T0) and those assigned a guaranteed outside option (T100). I use the average of the standardized test scores which are standardized by using the sample mean and standard deviation for the relevant test. "Any contribution" is a binary indicator if the job trainee engaged verbally ever during training. The "total number of contributions" is the cumulative number of contributions made by the job trainee during the whole training, and then separated out by quality as determined by the recruitment staff. "Late" is a binary indicator equal to 1 if the job trainee arrived late for training on that day. "Minutes early/late" is a continuous variable recording the minutes early (-) or late (+) job trainees arrived at training. Time use in columns 4 and 5 comes from survey data and is the number of hours conducting each activity daily. The effort index is a summary measure of the effort indicators. It is constructed as the average of the normalized values of: "Minutes early/late", " Hours studying training materials", "Hours watching television/listening to the radio".

Appendix Table B.8: Average effort indicators: Weighted results and bounds

	Punctuality			Hours studied training materials			Hours watching television or listening to the radio		
		Min-Max Bounds			Min-Max Bounds			Min-Max Bounds	
	Weighted	0-75=max; 100=min	0-75=min; 100=max	Weighted	0-75=max; 100=min	0-75=min; 100=max	Weighted	0-75=max; 100=min	0-75=min; 100=max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0% Job Guarantee	0.088	0.139	0.083	1.17	1.41	1.069	1.156	1.334	1.044
	[0.030]	[0.040]	[0.028]	[0.131]	[0.159]	[0.127]	[0.124]	[0.139]	[0.123]
1% Job Guarantee	0.081	0.091	0.079	1.158	1.268	1.127	1.593	1.681	1.536
	[0.025]	[0.027]	[0.024]	[0.111]	[0.134]	[0.110]	[0.134]	[0.146]	[0.134]
5% Job Guarantee	0.152	0.173	0.146	0.946	1.091	0.889	1.341	1.557	1.256
	[0.036]	[0.037]	[0.035]	[0.105]	[0.122]	[0.102]	[0.166]	[0.191]	[0.162]
50% Job Guarantee	0.079	0.125	0.073	1.087	1.222	1.03	1.505	1.658	1.429
	[0.030]	[0.040]	[0.029]	[0.100]	[0.121]	[0.099]	[0.133]	[0.150]	[0.136]
75% Job Guarantee	0.129	0.157	0.124	1.16	1.212	1.138	1.428	1.487	1.374
	[0.051]	[0.058]	[0.049]	[0.147]	[0.152]	[0.145]	[0.167]	[0.177]	[0.168]
100% Job Guarantee	0.186	0.182	0.186	0.742	0.73	0.747	2.029	2.014	2.037
	[0.066]	[0.067]	[0.066]	[0.078]	[0.090]	[0.074]	[0.247]	[0.246]	[0.250]
Observations	259	268	268	254	268	268	254	268	268
R-squared	0.24	0.27	0.24	0.69	0.65	0.66	0.71	0.69	0.68
p-values of F-test:									
0% and 100%	0.186	0.578	0.158	0.005	0.000	0.028	0.002	0.017	0.001

Notes:

This table presents the average daily effort by treatment group using both administrative data and survey data. "Late" is a binary indicator equal to 1 if the job trainee arrived late for training on that day. "Minutes early/late" is a continuous variable recording the minutes early (-) or late (+) job trainees arrived at training. Time use in columns 4 and 5 comes from survey data and is the number of hours conducting each activity daily. The effort index is a summary measure of the effort indicators. It is constructed as the average of the normalized values of: "Minutes early/late", " Hours studying training materials", "Hours watching television/listening to the radio". Treatment status was randomly allocated and stratified by quintile ability and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

Appendix Table B.9: Other Mechanisms: Weighted results and bounds

	Perceptions			**Food expenditures - groceries**			**Food expenditures - eat out**		
		Min-Max Bounds			Min-Max Bounds			Min-Max Bounds	
	Weighted	0-75=max; 100=min	0-75=min; 100=max	Weighted	0-75=max; 100=min	0-75=min; 100=max	Weighted	0-75=max; 100=min	0-75=min; 100=max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0% Job Guarantee	73.046	75.844	66.691	347.589	635.49	322.424	123.557	154.669	110.836
	[3.532]	[3.425]	[4.216]	[75.486]	[145.780]	[70.860]	[16.122]	[21.008]	[15.726]
1% Job Guarantee	73.541	74.3	70.995	425.016	576.75	407.061	165.349	179.587	158.686
	[2.992]	[2.914]	[3.494]	[97.883]	[149.297]	[96.436]	[15.036]	[17.657]	[15.313]
5% Job Guarantee	76.142	76.776	74.107	372.751	469.008	365.048	155.073	168.564	149.658
	[3.168]	[3.096]	[3.538]	[92.596]	[111.595]	[91.104]	[21.256]	[22.959]	[21.128]
50% Job Guarantee	72.651	74.246	70.518	438.595	665.541	416.322	147.121	183.593	138.344
	[2.339]	[2.419]	[2.586]	[97.284]	[158.985]	[93.549]	[19.970]	[28.294]	[19.617]
75% Job Guarantee	83.63	83.9	82.181	337.727	371.837	327.945	184.582	203.359	177.235
	[3.366]	[3.253]	[3.626]	[74.216]	[80.977]	[72.379]	[27.827]	[32.523]	[27.795]
100% Job Guarantee	77.596	77.543	77.902	328.642	309.028	329.545	123.838	119.957	124.523
	[3.562]	[3.649]	[3.405]	[79.859]	[89.147]	[79.408]	[23.189]	[22.202]	[23.549]
Observations	256	268	268	256	268	268	256	268	268
R-squared	0.94	0.94	0.91	0.36	0.31	0.35	0.6	0.57	0.57
p-values of F-test:									
0% and 100%	0.361	0.732	0.038	0.865	0.056	0.947	0.992	0.255	0.627

Notes:

This table presents the treatment group means for each outcome. Food Expenditures (in MKW) is the average amount spent on food reported by the respondent across the 3 training days. "Eat out expenditures (in MKW)" is similar except measures food expenditures for food consumed away from the home. "Perceived chance of employment with recruiter" is constructed using the following question: “What percentage chance do you think you have of getting one of the available positions for the RECRUITER’S PROJECT?” with the following options: No chance of getting a job; Less than 25 percent; Between 25 and 50 percent; 50 percent; Between 50 and 75 percent; Between 75 and 99 percent; and Certain about employment with recruiter. To create a measure of the likelihood of employment I assign the mid-point to categories that are brackets and creating a continuous variable.

Treatment status was randomly allocated and stratified by quintile ability and prior work experience with the recruiter. The stratification cell fixed effects include a set of dummies for each stratification cell. The set of additional covariates include: a dummy variable for whether the individual has worked before, marital status, age, and the individuals' standardized ability score. For covariates with missing observations the variable is assigned the mean value of the variable and an indicator variable is included for whether or not that particular variable is missing. Robust standard errors are presented.

APPENDIX C

Robustness checks II

This appendix presents additional specification checks for the analysis conducted in Chapter 2: Employment Exposure: Experimental Evidence on Employment and Wage Effects. Appendix Table C.1 provides additional information regarding differential attrition by treatment status and covariates.

Appendix Table A: Sample and Attrition				
	Baseline N=268			Covariate * Probability of
	Mean	SD	Covariate	Job offer
	(1)	(2)	(3)	(4)
<u>*Demographics:*</u>				
Age	25.604	4.638	0.004	0.001
Married	0.172	0.378	-0.031	0.136
Any child?	0.164	0.371	0.000	0.087
Number of children	0.299	0.784	0.028	-0.029
Years of education	13.183	0.940	0.064**	-0.104
Income (USD, 3 months)	206.123	228.803	0.00004	0.00001
Ability score	-0.001	1.003	0.035	-0.035
Tribe:				
Chewa	0.310	0.463	-0.064	0.093
Lomwe	0.108	0.311	0.125*	-0.304
Ngoni	0.164	0.371	0.057	0.138
Tumbuka	0.190	0.393	-0.041	0.112
Other	0.201	0.402	0.029	-0.188
<u>*Education and Work:*</u>				
Ever worked?	0.869	0.338	-0.014	-0.152
Ever worked with recruiter?	0.104	0.306	-0.093	0.107
Any work in last month	0.646	0.479	0.039	0.131
Any work in last 6 months	0.869	0.338	0.109	0.167
Frac of 6 mths worked	2.657	2.176	0.008	0.015
Any job search last month	0.116	0.320	-0.085	0.270**

<u>*Notes:*</u>

The baseline sample consists of 268 individuals who participated in the recruitment process and experiment discussed in Section 2. Columns 3 and 4 are from the same regression predicting where the dependent variable is whether or not the individual was found at follow up. Columns 3 and 4 present the coeffieint on the baseline characteristic and the interaction of the baseline coefficient and the assigned probability of a job offer respectively .