Teacher Quality and Student Learning: Evidence from the Gambian Hardship Allowance[1]

Todd Pugatch
Oregon State University and IZA

Elizabeth Schroeder
Oregon State University

December 2013

Abstract

We evaluate the impact of the Gambian hardship allowance, which provides a salary premium of 30-40% to primary school teachers in remote locations, on the enrollment, characteristics, and performance of students. A geographic discontinuity in the policy's implementation and the presence of common pre-treatment trends between hardship and non-hardship schools provide sources of identifying variation. Seven years after its implementation, we find no evidence that the hardship allowance increased enrollment, closed the gender enrollment gap, or improved learning for the average student. However, we do find that the hardship allowance increased enrollment of students from more advantaged households, and suggestive evidence of increased learning among high-ability students. The results suggest that an unconditional salary increase for teachers improves parents' perceptions of school quality, but only the best students benefit through improved learning.

[1] Department of Economics, School of Public Policy, Oregon State University, Corvallis OR 97331 USA. Contact: todd.pugatch@oregonstate.edu, liz.schroeder@oregonstate.edu.

Family Size and Human Capital in India

Santosh Kumar* Adriana Kugler†

First Draft: Sep 2011; This Draft: Feb 2012

Abstract

Using data from a representative sample from India, we test the empirical validity of Quantity-Quality trade-off model of Becker and Lewis (1973). To address the endogeneity arising from the joint determination of quantity and quality of children by parents, we instrument the family size by sex of the first child. We find a negative relationship between family size and children's educational attainment, even after controlling for parent's characteristics and birth order of children. The effects are heterogeneous. The trade-off is more pronounced in rural areas, for low-caste children, for illiterate mothers, and for children belonging to low wealth category. Overall, the findings support the quantity-quality trade-off in a resource poor country such as India. Given that for long-run economic development, the quality of human capital is equally important, policymakers should invest more in education and other welfare programs in order to mitigate the adverse impacts of larger family size.

JEL classification: N75, O18, O20.
Keywords: Quantity-Quality Trade-off, Family size, Education, India.

*University of Washington, Seattle, WA. (e-mail: skumar3@uw.edu)
†Georgetown Public Policy Institute, Georgetown University, Washington DC. (email: ak659@georgetown.edu). Please do not cite without author's prior permission.

1 Introduction

Policymakers in developing countries lay much emphasis on family planning programmes to reduce the family size in order to speed up the economic development. The policy intervention is based on the idea that a resource-constrained household with smaller family size will have more resources for investments in human capital of their children. In economics, this notion has been first modeled by Becker and Lewis (1973), which suggests that decrease in the quantity of children in a resource-constrained population will free up more resources to be invested on each child, leading to an increase in the average quality of child, quality often being measured by education or health status of children. In this paper, we test the empirical validity of children quantity-quality (Q-Q) model in India, instrumenting family size by sex of the first child.

Testing the Q-Q model in Indian context is interesting and is of tremendous policy relevance. Today, over 35% of Indian population is below the age of 20. By 2020, it is expected that 325 million people in India will reach working age, which will be the largest in the world. This will come at a time when the rest of the developed world will be faced with an aging population. It is estimated that by 2020, US will be short of 17 million people of working age, China by 10 million, Japan by 9 million and Russia by 6 million. At the same time, India will have a surplus of 47 million working people.[1]

The question that policymakers in India are facing today is whether this young Indian population is a bane or a boon. Economic growth requires not just a large working population (quantity), but people with high human capital, skilled enough to enter the labor market. If the quality of working population is not high enough to complement the sheer size of the working age population, the projected growth path may not be realized.[2] The working population size is a necessary but not sufficient condition for economic growth.

[1] Even when compared to developing countries, Brazil's working population is set to grow by 12%, China's by 1%, Russia's will decline by 18%, while India will grow by 30%.

[2] India's adult illiteracy levels are a big concern, which stands at 39%. 25 million children are out of school in India, out of a total of 100 million out of school children in the world.

Quality of population (such as education and health) is equally important to make them productive when they join the labor force. The population growth could "transform into a demographic dividend" only if every child is born healthy and is educated. To test this in the Becker and Lewis's Q-Q framework, we embark upon to examine the *causal* relationship between family size (quantity) and education attainment of children (quality) in India.

However, testing the existence of Q-Q trade-off empirically is challenging because child quantity and quality are endogenous variables, thereby confounding the causal interpretation. Fertility decisions and investment on child are jointly determined by parents (Browning, 1992; Haveman and Wolfe, 1995), which means that they are both affected by unobservable heterogeneity in parental preferences and household characteristics. Parents who are more concerned about the quality of their children may choose to have fewer children to educate each better.

The key method of addressing this endogeneity has been to take advantage of exogenous variation in family size that are either caused due to some policy experiment (one-child policy in China, forced sterilization in India), natural occurrences of twins birth, and sibling sex composition. The study by Rosenzweig and Wolpin (1980) is one of the early studies that exploits the birth of twins to isolate the causal effect of family size on child quality. The study found an inverse relationship between family size and children's educational attainment in a small sample of 25 twins in 1,600 children in India. In contrast, Qian (2009) exploits the exogenous variation in the family size due to relaxation of one-child policy norm in China, and surprisingly found a positive correlation between number of children and school enrollment. The author argues that a positive correlation is plausible due to economies of scale in raising the children.

In addition to twins and one-child policy, another plausible instruments used for family size are either gender of the first child or the gender composition of the first two children. The former instrument is based on the prevailing preference for sons that is observed in

Asian countries, and the idea behind the latter instrument is that parents of same-gender siblings are more likely to go on to have an additional child (more likely in developed countries). Using sex of the first child Lee (2008) finds negative impacts of family size on per-child investment in education for South Korean households. While Angrist el al. (2010) that used both twin births and gender composition of children as instruments found no evidence of a quantity-quality trade-off in Israel. Similarly, Millimet and Wang (2010) also fails to find statistically meaningful evidence of the children quantity-quality trade-off in Indonesia. The latter used gender composition of first two children as an instrument and examined the effect of number of children on child health outcomes such as BMI and height.

Following this strand of literature, this study uses sex of the first child to instrument the family size as measured by number of children. We believe that IV estimation method will yield consistent results under the reasonable assumption that this instrument is a random event. The preference for sons in East and South East Asian countries are widely documented. The son preferences are in countries like India, China, and Korea is deeply rooted in social, cultural, and economic factors. Preference for sons can alter the family configuration by affecting fertility. This is because the couple who have all the sons they want will stop bearing additional child, while couples who are hoping to for a son will continue having children. Therefore, in a son-biased society, the first child's sex should be a good predictor for the probability of having a second child or the total number of children (Lee 2008).

The empirical evidence on the validity of Q-Q model is mixed at least in developed countries. Black et al. (2005) shows that after controlling for birth order family size does not have significant effect on educational achievement of children in Norway. Angrist et al. (2010) explore the same question in Israel and do not find any evidence of quantity-quality trade-off. Similarly, Haan (2005) finds no significant effect of the number of children on educational attainment of the oldest child in US and the Netherlands. Caceres-Delpiano

(2006) finds a negative impact of family size on the likelihood that older children attend private school, but he finds no impact of an additional child on education attainment in US.

In contrast, using US PUMS data, Conley and Glauber (2005) show that children living in larger families are less likely to attend private school (attending private school is marker of high quality) and are more likely to be held back in school. Furthermore,Goux and Maurin (2005) show that children living in larger families perform worse in school than children in smaller families in France.[3] Additionally, Glick et al. (2007) use twinning at first birth to estimate the effects of unplanned fertility on the nutritional status and school enrolment of children in Romania. They find that a first-birth twins shock has negative impacts on children's human capital investments, particularly for later-born siblings.

Whether the absence[presence] of observed trade-off in developed countries also extends to developing countries is still an empirically open question, though in recent years, estimation of the trade-off has gained momentum in developing countries as well. This question has an important policy implication for developing countries since the resource constraint argument inherent in Q-Q model seems more logical in a resource poor setting. The earliest study on Q-Q trade-off in a developing country setting can be traced back to 1980, when using data from India between 1969 and 1971, Rosenzweig and Wolpin (1980) observed that educational attainment of children ws lower in high-fertility households. Recently a number of studies (Li et al., 2008; Qian, 2009; Rosenzweig and Zhang, 2009) have found mixed evidence of the Q-Q trade-off in China. Qian (2009) finds positive effect, while Li et al. (2008) finds some evidence of a Q-Q trade-off in rural areas of China but not in urban areas. Rosenzweig and Zhang (2009) finds that an extra child significantly decreases the schooling progress, the expected college enrolment, grades in school and the assessed health of all children in the family.

[3]Using marital fecundability- as measured by the time interval from the marriage to the first birth- as a source of exogenous variation in family size, Klemp and Weisdorf (2011) documents a large and significantly negative effect of family size on children's literacy in UK.

Lee (2008) shows that per-child investment in education in South Korea is less in households which has larger family size. Using twinning as an instrument, Ponczek and Souzay (2011) also finds negative effects on educational outcomes in Brazil, while Aguero and Marks (2008) find evidence of Q-Q trade-off only for health indicators (weight-for-age and breastfeeding practice) but not for education.[4] Similarly, Razak et al. (2010) found no evidence in support of Q-Q trade-off in Malaysia.

Our paper is related to scant literature that uses sex of first birth (henceforth, SFB) to estimate the causal effects of family size on child education. As in this paper, Lee (2008) also uses sex of first birth as an instrument for the family size in South Korea. Using a representative sample from India, we find evidence in support of the Q-Q trade-off. Our results show a negative effect of larger family size on educational attainment of children. In the 2SLS estimation, an extra child in the family reduces the probability of primary school completion by 5 percentage points and years of schooling by 0.36 years. Moreover, we find that the negative effects on educational outcomes are more pronounced for children in rural and low-caste households. We also find the trade-off is more severe for illiterate mother and low-wealth households.

Our paper adds to the existing literature on quantity-quality trade-off in an important ways. Most importantly, compared to most of the previous studies, this paper looks at a poor country, where extent of trade-off can be acute due to scarcity of resources.[5] We are the first to test the validity of Q-Q trade-off in India where almost one-sixth of world's population reside.[6] Second, it is quite likely that the extent and nature of trade-off among Indian households are different compared to households in developed and other developing countries.

The characteristics of credit markets could affect the extent and severity of quantity-

[4]The study uses data from Demographic and Health Surveys in Latin America, namely, Bolivia (conducted in 1994 and 1998), Brazil (1996), Colombia (1995 and 2000), the Dominican Republic (1996), Guatemala (1998), Nicaragua (1998), and Peru (1996), and instrument the family size by a mother's infertility status.

[5]Qian (2009), Rosenzweig and Zhang (2009) and Li, Zhang, and Zhu (2008) focus on China while Lee (2008) focuses on South Korea.

[6]Although, the study by Rosenzweig and Wolpin (1980) is the first paper on Q-Q in India, however due to very small sample size, results can not be assumed to be true in other parts of India given the heterogeneity in Indian culture and institutions.

quality trade-off. For example, in a setting where credit markets are not well-developed, are imperfect and households are unable to smooth out consumption, the extent of Q-Q trade-off could be more acute and severe. Similarly, households who are more credit constrained could face a very high level of trade-off compared to households who are not credit-constrained. Estimating the effect of family size on child educational outcomes in India is important from policy perspective as majority of Indian households are poor, credit constrained, and have large family size.

The paper is organized as follows. The next section discuss the instrument and its validity. Section 3 explains the empirical framework followed by Section 4 that introduces the District Level Health Surveys (DLHS) and discuss the data used in this paper. The main results are presented in section 5, while section 6 presents the heterogeneity in the results. Finally, Section 7 closes the paper by summarizing the findings and outlining the agenda for future research and policy considerations.

2 Empirical Framework

Using sex of the first birth (SFB) as an instrument, we investigate the effect of family size on children's educational outcomes. We employ OLS and 2SLS regression analyses on the sample of 311,942 children as described in Table 1 below. Formally, we estimate the following OLS equation:

$$EDU = \beta_0 + \beta_1 * FamilySize_i + \beta_2 X_i + \mu_d + \epsilon_i \tag{1}$$

Where, EDU is the educational attainment of the child as measured by probability of being literate, probability of ever attended school, probability of currently being enrolled, grade progression, years of schooling, and probability of primary school completion.[7] The variable *FamilySize* is the number of 0-20 years old children in the family at the time of

[7] Grade progression= (years of schooling)/(age-5). Primary school completion outcome is analyzed for 14-20 years children in the sample while other outcomes are analyzed for 0-20 years old sample.

survey; X is a vector of covariates, and ϵ is an error term. Covariates comprise of child characteristics, including age, gender, ethnic group, birth order and place of residence (rural vs urban). In addition to these child-level characteristics, we also include a set of parental attributes, including age and education level of children's father and mother. The main coefficient of interest is β_1 that will provide the evidence on Q-Q trade-off. A negative β_1 would mean that quantity-quality trade-off does exist in Indian society.

As already pointed out, since fertility behavior and decisions about childrenŠs education are jointly determined, the OLS estimates of equations (3) are subject to endogeneity bias and do not necessarily depict the causal effects. The OLS estimates will be either downward or upward biased depending on nature of endogeneity. For example, a pure income effect would make such omitted factors positively related to both fertility and education; in this case, the OLS coefficients would underestimate the effect of an additional child on children's quality.

However, it can also be imagined that parents' tastes and preferences may affect fertility negatively (they are better able to understand the use of birth-control policies). If these preferences and tastes are also positively correlated with children's educational outcomes, the OLS estimates will overestimate the true causal impacts of family size on children's quality.

To address these potential concerns about endoegeneity, we use the two stage least square (2SLS) regression in our analysis. We estimate the following two-stage least square model:

$$FamilySize = \alpha_0 + \alpha_1 * SFB_i + \alpha_2 X_i + \mu_d + \epsilon_i \tag{2}$$

$$EDU = \alpha_0 + \alpha_1 * \widetilde{FamilySize}_i + \alpha_2 X_i + \mu_d + \epsilon_i \tag{3}$$

where *FamilySize* is number of children as measured by household size, *EDU* is mea-

sured by primary school completion and completed years of schooling, *SFB* is the instrumental variable for *FamilySize*, and X denotes vector of exogenous regressors. SFB is a dummy variable that equals 1 if the first-born is a female and 0 otherwise. We also include district fixed-effects to control for time-invariant differences across districts. Standard errors are clustered at district level.

Equation (2) is the first-stage while equation (3) estimates the second-stage results. The second stage regress the outcomes on the predicted value of family size from equation (2) and other exogenous variables.

3 Data and Sample Statistics

The data used in this study are taken from the third round of Indian District Level Household Survey (DLHS) collected in 2007-08. The sample is representative at district level, the lowest level of administration and policy-making. The DLHS covers 601 districts and on average draws a random sample of 1000-1500 households in each districts.

The survey has three modules: household roster, ever-married women (15-49 years), and unmarried women (15-24 years) in addition to village and health facilities survey. The household questionnaire collected information on all members of the household and socio-economic characteristics of the household, assets possessed, number of marriages and deaths in the household since January 2004. The ever-married womenŠs questionnaire contained information on womenŠs characteristics, maternal care, immunization and childcare, contraception and fertility preferences, reproductive health including knowledge about RTI/STI and HIV/AIDS.

We used household module to construct our analytical sample. We were unable to use ever-married women's module as the survey did not gathered information on the complete fertility history of the women unlike the standard Demographic Health Survey (DHS). The DLHS-3 only collected fertility history of women for children born since 1 January, 2004.

Therefor, we rely on information in household roster to construct our sample for analysis. Specifically, we use the information on relation to the household head in the household roster to construct our analytical sample. The household roster contains extensive information on personal and household characteristics. For each person in the household, information about, e.g., age, gender, schooling attendance, literacy, years of completed schooling, is available. We identify individuals who are labeled "sons/daughters" as the primary observation, and then obtain the family size by counting the number of children in the household. We then attach the data of parents, those who are labeled as "household head" or "spouse" to all the sons & daughters in the household.

For simplification, we trim the sample in the following ways. First, we restrict the sample to individuals who are either parents (head of the household, and spouse) or own children (who are either sons/daughters of the head of the household).[8] Second, we restrict the sample to households with atleast one child so that we can use gender of the first child as the instrument. Third, we restrict the sample to young mother who are not older than 35 at the time of survey, and the children to school going age of 5-20 years. We use 5 as the lower age bound because the household roster only collected education information of all individuals older than 4. In India, primary school (grades 1 to 5) begins at age 5 or 6 and ends at age 10 or 11, while high school is usually complete by age 18. However, due to deferred enrollment or grade repetition completion of either primary or secondary schooling might get delayed. Finally, we keep households whose eldest child is less than 21 years old.

We exclude mothers over 35 years of age to minimize the possibility of migration of adult children out of the household. Finally, we exclude households with missing or unreliable information on any of the variables used in the analysis. This gives us an analytical sample of 332,689 children.

The main outcome variables this paper analyzes are different measures od educational

[8] We drop individuals who are son or daughter-in-law, grandchildren, parent, parent-in-law, brother/sister, brother or sister-in-law, niece or nephew, other relative etc.

attainment of 5-20 years old children. The outcome measures are probability of being literate, probability of ever attending school, probability of currently being enrolled, grade progression, years of schooling, and probability of primary school completion. Primary school completion outcome is analyzed for 14-20 years children in the sample because by age 14, primary school should be complete.

In all the models stated above, we consider the following variables as an additional covariates: caste, religion, an asset-based standard of living index, mother's age, father's age, mother's education, father's education. For the caste variable, we consider three groups: scheduled caste and scheduled tribe are combined together to constitute the low caste category (a group that is socially segregated and disadvantaged), and other backward classes (officially identified as socially and educationally backward) as middle caste, and the upper caste (comprising Brahmins and other higher castes that are privileged) as high caste. We consider 4 major religious groups, Hindus, Muslims, Sikhs and Christians. The DLHS data does not contain information on individual or household incomes. The survey asked a multitude of questions about the ownership of assets like a car, television, property etc. The DLHS has used ownership of assets to create a standard of living index (SLI) with three categories: low, middle and high.

Table 1 reports the summary statistics of these individual and household predictor variables. The average age of children in the sample is 10 years with 3.07 as average years of schooling. About 52 percent of first born children are male. Around 76 percent of children in 14-20 age group have completed primary schooling. Fathers are relatively older than mothers. The average age of mothers is 31 years while fathers are 37 years old. We kept young parents in the sample to make it fairly certain that no adult children have moved out of a household. We impose such a restriction because the survey does not contain information about children who had already left the household by the time of the survey. In spite of this restriction, we cannot fully rule out that no adult children have moved out. As expected, mothers have less education than fathers. The average years of

schooling for mothers is 3.1 years, while on average fathers have completed 5.7 years of schooling.

The average family size is 3.55 with majority of the children residing in rural areas. In the sample, approximately 82 percent of children lives in rural areas. About two-fifth of the children belongs to low caste and one-fifth to high caste. Finally, by the standard of living index, around 49% of children belong to the group with lowest standard of living index, 39% belong to middle-wealth category; while the rest 12% belong to high-wealth category.

[TABLE 1]

4 Son Preference and Instrument Validity

In this section, we explain why Indian parents prefer sons over daughters and whether sex of first birth can really be assumed a random event and whether it can be treated as natural experiment. A related question we explore in this section is "Is the gender of first-birth a valid instrument for family size?" In order for the sex of first-birth to be a good instrument it (1) has to be correlated with family size and (relevance) (2) cannot directly affect educational outcomes (exclusion restriction).

The question of whether the exclusion restriction is satisfied translates into a question of whether sex of first birth (henceforth SFB) is indeed exogenous, and whether it is correlated with any omitted variables which would affect educational outcomes of children. In this section, we discuss the validity of sex of first child as the instrument. The instrument, sex of the first child, is a valid instrument if it satisfies the following two conditions:

$$Relevance : Corr(Z_i, X_i) \neq 0 \tag{4}$$

$$Exogeneity : Corr(Z_i, \epsilon_i) = 0 \tag{5}$$

Condition (1) implies that the instrument, *SFB (Z)*, should be highly correlated with the endogeneous variable, *family size (X)*, and condition (2) implies that the instrument should not affect the child outcomes except through family size. In a country like India where son preferences persist, the sex of first child is an important source of exogenous variation in fertility. The sex of the first birth has been used in the previous research by Lee (2008) in Korea. Another study by Li and Wu (2010) have also used sex of first born as an instrument to estimate the causal impact o women's bargaining power in China.

In India, there is a long-standing social and cultural norm of sons preference over daughters because only sons could carry the family name and inherit the family patrimony. India is a patriarchal society. Parents prefer sons as sons are supposed to provide financial support and care in old ages. As per Indian tradition, daughters are married out and become part of another family. Economic reasons also contribute to the practice of son preference as they are more likely to join labor markets and are more productive. In this type of patrilineal familial system the gender of first-born has an important implication for the family size and that motivates our reliance on gender of the first child as an instrument.

However, the existence of sex-selective abortion may undermine the validity of the instrument because the access to ultrasound use and abortion services allows parents to choose the sex of their children. However, this does not seem to be a big concern given that ultrasound technology is not widely available in rural areas and more importantly, passing of Pre-natal Diagnostic Technique (PNDT) Act in 1996 made the fetal-sex determination illegal.

As stated above, validity of our instrument relies on the fact that the sex outcome of a first-birth in India is a random event and plausibly exogenous. Many previous studies have shown that parents do not use sex-selective abortion for their *first pregnancy*- they argue that sex-ratio at first birth lies within the biologically range of 1.03-1.07 (Bhalotra and Cochrane (2010), Ebenstein (2007), Jha et al. (2011), Portner (2010), and Rosen-

blum (2010)).[9] Sociological studies also provide evidence that parents only have a strong preference for sons after this first birth (Patel, 2007).

Using the same data as ours, Rosenblum (2010) reports lack of sex-selection abortion at the first-parity. The paper reports lack of induced abortion by women for the first-parity. Incidentally, about 36 percent of women report induced abortions at the second and third-parities. Given that sex-selection abortion became illegal in India in 1996, zero and positive reporting of induced abortion at first and high-order respectively further provides confidence that sex-selection at first parity is not rampant and gender of first-born can really be treated as exogenous. Additionally, Retherford and Roy (2003) using the first two rounds of the National Family and Health Survey finds little or no evidence of sex selection at the first birth. Taken together, these studies provide overwhelming evidence that sex of first birth is indeed exogenous and random.

In order to further examine whether sex of first birth is truly exogenous, we estimate a linear probability model of sex of the first birth on a vector of explanatory variables. Additionally, we also estimate a probit model. Table 3 presents the results. It can be seen in Table 2 that none of the explanatory variables except mother's age are statistically significant. These results provide additional evidence in support of sex of first birth being truly random as none of the background characteristics are able to explain the gender of he first child. Furthermore, in our analytical sample, about 52% of first-born are male indicating that the sex-ratio at first birth is in the biological range (Table 1).

[Table 2]

5 Main Results

The outcome variables in this study include a set of educational indicators. These include: (i) literacy; (ii) ever attended school; (iii) school enrollment; (iv) grade progres-

[9] In the absence of any interventions the probability of having a son is approximately 0.512 and this probability is independent of genetic factors (Ben-Porath and Welch 1976; Jacobsen, Mïller and Mouritsen 1999).

sion; (v) years of schooling for those who ever attended school; and (vi) primary school completion for 14-20 years old sample. Except grade progression and years of completed schooling, all other variables are binary and are coded as 1 and 0. The main independent variable, household size, is continuous and is measured by the total number of 0-20 years old children in the family at the time of survey. The OLS regression results are presented in Table 3.

All the models in Table 3 control for child and parent's characteristics. Additionally, district fixed-effects are also included to account for district fixed characteristics. We estimate Linear Probability Model (LPM) for the binary outcomes (i, ii, iii, and vi), while for grade progression and years of completed schooling, OLS regression models are estimated.[10] Results in Table 3 indicate a significantly negative correlation between family size and children's education. Results suggest that after controlling for children's characteristics and other confounding covariates, on average, an extra child in the family reduces the probability of literacy by 2 percentage points (col 1). Similar negative coefficients are also observed for other educational measures of children's quality.

Children in large families are 1.7 percentage points less likely to have ever attended school (col 2) and probability of being currently enrolled in school is also lower by 0.9 percentage points (col 3). The results for primary school completion, grade progression, and years of schooling also indicate harmful effects of family size on children's quality. An extra child in the family reduces the probability of primary school completion by 3.5 percentage points, while grade progression reduces by 0.03. Finally, for years of schooling, the point estimate is -0.156, suggesting that on average children in large families attain 0.16 fewer years of schooling.

[Insert Table 3]

Recognizing the limitation of interpreting the OLS estimates in Table 3 as causal, we

[10] Columns (1)-(3) estimate linear probability model instead of logit or probit models. The main difference between the two regressions is that the linear probability model assumes marginal effects to be constant, whereas the logit/probit model allows marginal effects to change. Close to the average values of the regressors, the linear probability model is a good estimator.

then proceed to instrument the main endogenous variable, family size, by sex of the first birth, and estimate the same relationship in 2SLS framework. For the 2SLS estimates to be meaningful, the instrument, *SFB*, should be correlated with the endogenous variable family size (relevance condition), but should not affect the educational outcomes directly, rather the instrument should affect the outcomes indirectly through the endogenous variable (exclusion condition).

We check the relevance condition in Table 5. From the first-stage regression, it follows that the instrument is highly significant and has a positive effect on family size. The birth of first child as female increases the family size by 0.25 children (Column 1 of Table 5) and the effect is significant at 1% level of significance. Other coefficients of the instrument are also positive and statistically significant. Table 5 also reports the Kleibergen-Paap rk Wald test to detect if the instrument suffers from weak-IV problem. The first-stage F-stat and Kleibergen-Paap rk Wald Stat, both are significant, suggesting that our analysis does not suffer from weak identification. We also provide Anderson-Rubin F-test Statistic and Stock-Wright S-statistic in Table 5 to assess that our second-stage results are robust to weak-instrument inference.

The first stage coefficient is smaller than the twins first stage of about 0.6 in the Angrist and Evans (1998). Birth of first child as girl results in a smaller increase in family size due to the fact that Indian families normally desire to have larger family size.

The IV results show a negative and significant impact of family size on children's quality except that the result on primary school completion is not significant and infact the point estimate is positive. The IV estimates for literacy is 2.6 percentage points, which corresponds to a decrease of 3 percent (0.026/0.86) for the probability of being literate. Similarly, the point estimates for ever being in school and current enrollment are negative and statistically significant, suggesting that the detrimental effects of family size on children's education.

In Bangladesh, the official age of entry at Grade 1 is six, it is expected that children

will be promoted to Grade 2 in the next year at the age of seven and to the next year to Grade 3 at the age of eight and so on. However, for various reasons this is not happening.

The Iv results on grade progression and years of completed schooling are again negative and substantial. For example, the coefficient for grade progression is -0.018. Since the average grade progression is 0.67, this means that an additional child in the family would decrease the grade progression by 0.12

The first two columns in Table 3 report the coefficients for primary school completion as the dependent variable, while the last two columns report the coefficients for years of schooling. The 2SLS estimates confirm the negative correlation observed in OLS estimation. We see that, once we account for the endogeneity of family size (using IVs), the coefficients are much larger in magnitude as compared to the ordinary least square estimate. The 2SLS coefficients are statistically significant at 1% level of significance and are almost three times bigger than the OLS estimates. The IV estimates suggest that OLS coefficients underestimate the true trade-off and are downward biased. According to 2SLS estimates, each additional child reduces the probability of completing primary school among all children in the family by 5 percentage points and schooling by 0.38 years.

[Insert Table 5]

6 Heterogeneity in the Trade-off

6.1 Caste and Rural-Urban Differences

Given that there is considerable rural-urban gap in the family size and educational attainment in India, we hypothesize the effect of family size on children's education to be different in rural and urban areas. For example, for our sample children, the primary school completion rate is 35% in rural areas while it is 41% in urban areas. In order to capture the heterogeneity across different caste categories, we divide the sample into different caste groups. We compare the effect of family size on education across low-vs-

middle-high caste children. The results are presented in Table 4. The upper panel in Table 4 shows the effect of family size on primary school completion while the lower panel displays the results for years of schooling. The results for caste categories in the first three columns show negative effects on primary school completion of 5.8 percentage points for low-caste children and of 5.1 percentage points for high-caste children. Surprisingly, the impact of of having larger family size is larger and statistically significant in urban areas compared to rural areas, suggesting that quantity-quality trade-off is large in urban areas.

[Insert Table 4]

The lower panel in Table 4 presents results for years of schooling. The coefficient of having a larger family size is much larger and more significant for low-caste and rural families, suggesting much severe trade-off in these families. The 2SLS estimate for low-caste children is 0.44 years and 0.31 years for high-caste children. We find caste gradient in the trade-off- the effect size decreases in magnitude as the children move into higher caste category. We observe a similar pattern in rural areas. The trade-off coefficient is 0.41 and highly significant in the rural areas while in the urban sample the trade-off coefficient is insignificant, suggesting that quantity-quality trade-off exists mainly in rural areas. This finding is similar to Li, Zhang, and Zhu (2008), who also found that trade-off was more evident in rural parts of China and was negligible in urban areas.

6.2 Mother's Education and Wealth Differences

The degree of trade-off might also be different across household wealth levels and mother's education levels. In this section, we allow the effect to vary by mother's education levels and household's wealth levels. We divide the sample by mother's education as illiterate, less than primary, and primary and more. We estimate the main model on these three samples separately. Similarly, we also divide the sample by household wealth levels to explore whether the trade-off differs among low, median, and high-wealth levels and run the IV model separately on these categories. Results on the stratified sample

are presented in Table 5. In the first three columns, sample is broken down by mother's education and the last three columns reports results by wealth levels.

The coefficients across OLS and 2SLS models are consistent with our expectations. We expect to see a larger effect for illiterate and less-educated mothers due to lack of financial resources. Based on the same argument, we also expect to see a larger coefficient for the trade-off in less well-off households. In the 2SLS model, the trade-off coefficient decreases in size with the wealth-level for primary school completion as well as years of schooling. However, for the high-wealth families, the 2SLS coefficients are statistically insignificant.

[Insert Table 5]

To a large extent, the story remains the same when we stratify the sample by mother's education level. For primary school outcome, there is a negative effect of family size for illiterate and primary school completed mothers. For illiterate mothers, the trade-off coefficient is very severe- arrival of an extra child reduces the probability of completing primary school by 8.7 percentage points and reduces years of schooling by 0.7 years. We do not find any evidence of trade-off in less than primary and primary completed mothers when the educational outcome analyzed is years of schooling. Though the coefficients monotonically decreases with the level of mother's education, the estimates are not significant in col 2 and 3.

7 Conclusions and Discussions

Testing the theoretical trade-off between the quantity and quality of children has been on the research agenda for a long time, however, the empirical evidence supporting the prediction of Beckerian model is limited. The empirical evidence has been mixed so far. A few studies find a negative effect of family size on the quality of children, measured by either education or health status (Ref). In contrast, others find no empirical support for the child quantity-quality trade-off (Ref). A variety of instruments such as twinning, sex

of first child, sex of first two child, infertility etc. are used to address the endogeneity concern.

In this paper, we have used household data from India to test the empirical validity of child quantity-quality trade-off. A strong preference for sons over daughters in Indian societies allows us the use of a novel instrumental variable, namely sex of first birth, to test the Q-Q trade-off. Testing this model has important policy implications. From policy point of view, it is important to know the extent to which a policy formulated to control population improves the human capital of the country and quality of the labor force. Not only the quantity of human capital plays a role, rather quality of human capital is equally important for economic development.

We find that Beckerian theory of child quantity-quality trade-off holds in India. Family size has significant negative causal impact on educational outcomes of children. After controlling for potential endogeneity, an additional child in the family reduces the probability of completing primary school for all children by 5 percentage points and years of schooling by 0.36 years, hence a strong support to Becker's trade-off hypothesis. The observed trade-off exists after including child and parents characteristics. We find non-uniformity in the existence of trade-off between rural and urban India. The negative relationship between family size and children's education is more pronounced and evident among rural households who are severely budget-constrained. Urban children are less likely to face the trade-off.

The effect also differs by caste, mother's education level, and household wealth. For children belonging to low and middle caste, the trade-off is severe compared to high-caste children. More educated mothers are also able to mitigate the trade-off as the trade-off is only evident for illiterate mothers. Similarly, we observe a wealth-gradient in the trade-off across wealth groups, with trade-off being more pronounced in low-wealth households with an extra children reducing the years of schooling by as high as 0.6 years.

Our findings are also supportive of theoretical work by Galor and Moav (2002), who

were the first to argue that the quantity-quality trade-off was decisive to economic advancement, not just from the onset of the demographic transition, but throughout human history.

Correctly estimating the causal effect of family size on child-quality outcomes is important for a developing country's public policy perspective. The majority of large families are poor, and our results suggest that family size has a direct impact on important outcomes for children. This discussion can better inform the public debate about how to understand and address poverty, education, and child labor in developing countries.

Our results suggest that policymakers in developing countries should invest more in education in areas and households for whom the trade-off is severe in order to mitigate the adverse impacts of larger family size.

References

Angrist Joshua and Williams Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *American Economic Review* 88(3):450-477.

Angrist Joshua, Victor Lavy, and Analia Schlosser. 2005. "New Evidence on the Causal Link Between the Quantity and Quality of Children." NBER Working Papers 11835, Cambridge: National Bureau of Economic Research.

Angrist Joshua, Vicotr Lavy, and Analia Schlosser. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labour Economics*, 28:773-824.

Black, Sandra, Paul Devereux and Kjell Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *Quarterly Journal of Economics*, 120(2):669-700.

Bhalotra, Sonia and Tom Cochrane. (2010). "Where Have All the Young Girls Gone? Identification of Sex Selection in India." IZA Discussion Paper, 2010, 5381.

Caceres-Delpiano, Julio. 2006. "The Impacts of Family Size on Investment in Child Quality." *Journal of Human Resources*, XLI(4):738-754.

Conley D., Glauber R. (2005) Parental Educational Investment and Children's Academic Risk: Estimates of the Effects of Sibship Size and Birth Order from Exogenous Variation in Fertility. New York: University of New York.

Galor, O. and O. Moav (2002.) "Natural selection and the origin of economic growth,"*Quarterly Journal of Economics*, 117:1133-1191.

Glick, P., A. Marini, and D.E. Sahn (2007) Estimating the Consequences of Unintended Fertility for Child Health and Education in Romania: An Analysis Using Twins

Data. *Oxford Bulletin of Economics & Statistics*, 69, 667-691.

Goux, Dominique and Eric Maurin (2005) The effect of overcrowded housing on children's performance at school," *Journal of Public Economics*, 89(5-6), 797-819.

Haan, Monique De (2005) , Birth Order, Family Size and Educational Attainment." Tinbergen Institute Discussion Papers 05-116/3, Tinbergen Institute December 2005. available at http://ideas.repec.org/p/dgr/uvatin/20050116.html.

Ebenstein, Avraham. 2007. "Fertility Decision and Sex Selection in Asia: Analysis and Policy," Mimeo.

Jha, Prabhat et al. 2011. "Trends in selective abortions in India: Analysis of nationally representative birth histories from 1990 to 2005 and census data from 1991 to 2011," Lancet, 377 (9781), 1921-1928.

Klemp P.B., Mark and Weisdorf L. Jacob (2011) The Child Quantity-Quality Trade-Off during the Industrial Revolution in England, Working Paper, department of Economics, University of Copenhagen.

Lee, Jungmin (2008) "Sibling size and investment in children's education: an asian instrument", *Journal of Population Economics*, vol. 21, issue 4, pages 855-875.

Li, Hongbin, Zhang, Junsen, and Zhu Yi (2008) "The Quantity-Quality Trade-Off of Children in a Developing Country: Identification using Chinese Twins", *Demography*, vol. 45, number 1, pages 223-243.

Patel, Tulsi, ed. Sex-Selective Abortion in India, New Delhi, India: Sage Publications,2007.

Portner, Claus. [2010. "Sex Selective Abortions, Fertility and Birth Spacing," University of Washington, Department of Economics, Working Paper, 2010, UWEC-2010-4.

Ponczek, Vladimir, Souzay P. Andre (2011) New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country, C-Micro Working Paper Series, Sao

Paulo School of Economics, Getulio Vargas Foundation.

Qian, Nancy (2009) Quantity-Quality and the One Child Policy: The Only Child Disadvantage in School Enrollment in China. Working Paper, Department of Economics, Yale university.

Retherford, Robert D and T K Roy (2003) Factors Affecting Sex-Selective Abortion in India and 17 Major States, National Family Health Survey Subject Reports 21, International Institute for Population Sciences, Mumbai, India.

Rosenzweig, M.R. and K.I. Wolpin (1980) Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment. *Econometrica*, 48(1):227-240.

Rosenzweig, M.R. and K.I. Wolpin (2000) Natural "Natural Experiments" in Economics. *Journal of Economic Literature*, 38, 827-874.

Rosenzweig, M.R. and J. Zhang (2009) Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's 'One Child' Policy. *Review of Economic Studies*, 76, 1149-1174.

Wang, Lee, Millimet, Daniel (*Forthcoming*) "Is the Quantity-Quality Trade-off a Trade-off for All, None, or Some?" *Economic Development & Cultural Change*.

Table 1: Descriptive Statistics of the sample

	Mean	Standard Deviation
	(1)	(2)
Age (6-20 years old)	10.32	3.26
Gender of first child (male=1)	0.52	0.47
Literate	0.86	0.34
Ever attended school	0.90	0.30
Still enrolled	0.95	0.21
Grade progression	0.67	0.56
Years of schooling	3.07	2.91
Primary school	0.76	0.44
Mother's age	30.93	3.36
Father's age	36.99	4.71
Mother's years of schooling	3.11	4.07
Father's years of schooling	5.68	4.70
Family size	3.55	1.33
Rural	0.82	0.39
Low caste (SC & ST)	0.41	0.49
Middle caste (OBC)	0.39	0.49
High caste	0.20	0.40
Low wealth	0.49	0.50
Median wealth	0.39	0.49
High wealth	0.12	0.33
No of observation	393,597	
No of district	601	

Notes: Standard deviations are shown in parentheses. All sampled children were 6-20 years old at the time of survey (2007-08) Mother's sample is restricted to 20-35 years.

Table 2: Regression of Gender of First Born on Control Variables

	Dependent variable: First-born is a girl	
	LPM	Probit
	(1)	(2)
drural	-0.003	-0.008
	(0.004)	(0.011)
poor	-0.006	-0.015
	(0.006)	(0.016)
medium	-0.004	-0.009
	(0.005)	(0.013)
hindu	0.007	0.017
	(0.004)	(0.011)
dscst	0.003	0.008
	(0.004)	(0.011)
dobc	0.002	0.005
	(0.004)	(0.010)
dm_ill	-0.008	-0.020
	(0.004)	(0.011)
dm_prim	-0.003	-0.007
	(0.004)	(0.011)
df_ill	-0.009*	-0.024*
	(0.004)	(0.010)
df_prim	-0.011**	-0.028**
	(0.004)	(0.010)
mage1	0.037***	0.093***
	(0.007)	(0.016)
mage2	-0.001***	-0.002***
	(0.000)	(0.000)
fage1	0.002	0.005
	(0.003)	(0.008)
fage2	-0.000	-0.000
	(0.000)	(0.000)
District dummies	Yes	Yes
r2	0.007	

Notes: Standard deviations are shown in parentheses. Column 2 reports the marginal effects in a probit model.
* Significant at 10 percent; ** significant at 5 percent
*** significant at 1 percent

Table 3: OLS estimates of the effect of family size on education- Pooled; 2+

	Literacy	Ever attended school	Currently enrolled	Grade progression	Years of schooling	Schooling (≥ 5)
	(1)	(2)	(3)	(4)	(5)	
hhsize	-0.020***	-0.017***	-0.009***	-0.017***	-0.156***	-0.035***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.005)	(0.002)
Children's control	yes	yes	yes	yes	yes	yes
Parents' controls	yes	yes	yes	yes	yes	yes
District fixed-effect	yes	yes	yes	yes	yes	yes
No of observations	393597	393597	345985	258965	393597	82640

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, gender, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of children in the family at the time of survey. Instrument is the a dummy variable indicating if the first-born is a female child. The regressions in columns 2-4 are calculated on the sub-sample from column 1 that took the learning test.

Table 4: First Stage

	Number of children			
	(1)	(2)	(3)	(4)
First girl (FG)	0.267***	0.234***	0.228***	0.204***
	(0.017)	(0.016)	(0.012)	(0.017)
fdscst	-0.041*			
	(0.021)			
fdobc	0.004			
	(0.020)			
fpoor		-0.014		
		(0.019)		
fmedium		0.063***		
		(0.018)		
fdm_ill			0.021	
			(0.016)	
fdm_prim			0.072***	
			(0.020)	
fdrural				0.058**
				(0.019)

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error,clustered by district, are shown in parentheses.

Table 5: IV estimates of the effect of family size on children's educational outcome: 2+ sample

	Instrument: First child is a girl (G)					
	Lit	Ever attended school	Currently enrolled	Years of schooling	Grade Progression	Schooling$\geq$ 5 age (12-18)
Family size	-0.027 ***	-0.018***	-0.009***	-0.069***	-0.010	-0.009
	(0.006)	(0.005)	(0.003)	(0.029)	(0.009)	(0.014)
First Stage	0.256***	0.256***	0.271***	0.253***	0.255***	0.252***
	(0.009)	(0.009)	(0.009)	(0.012)	(0.009)	(0.008)
Weak-Identification Tests						
Kleibergen-Paap Wald rk F-stat	814.61	814.61	932.80	421.51	851.93	887.92
p-value	0.00	0.00	0.00	0.00	0.00	0.00
Weak-Instrument-Robust Inference						
Anderson-Rudin F	11.77	11.77	5.86	0.02	4.29	5.29
p-value	0.0006	0.00	0.0158	0.88	0.03	0.02
Stock-Wright S stat	11.36	11.36	5.79	0.02	4.23	5.18
p-value	0.0008	0.00	0.0161	0.88	0.03	0.02
Children's control	yes	yes	yes	yes	yes	yes
Parents' controls	yes	yes	yes	yes	yes	yes
District fixed-effect	yes	yes	yes	yes	yes	yes
N	393597	393597	345985	393597	258965	82640

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of children in the family at the time of survey.

Table 6: IV estimates of the effect of family size on children's educational outcome: 3+ sample

	Instrument: First two children are girl (GG)					
	Lit	Ever attended school	Currently enrolled	Years of schooling	Grade Progression	Schooling $\geq$ 5 age (12-18)
	(1)	(2)	(3)	(4)	(5)	(6)
Family size	-0.015 ***	-0.018***	-0.004	-0.159***	-0.016**	-0.019**
	(0.005)	(0.004)	(0.003)	(0.022)	(0.007)	(0.010)
Children's control	yes	yes	yes	yes	yes	yes
Parents' controls	yes	yes	yes	yes	yes	yes
District fixed-effect	yes	yes	yes	yes	yes	yes
N	300609	300609	260702	300609	197364	65054

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of children in the family at the time of survey.

Table 7: 2SLS estimates of the effect of family size on education by caste and residence; 2+

	Low Caste	Middle Caste	High Caste	Rural	Urban
	(1)	(2)	(3)	(4)	(5)
Lit	-0.043***	-0.020*	-0.008	-0.029***	-0.021
	(0.011)	(0.010)	(0.010)	(0.007)	(0.014)
EverSchool	-0.033***	-0.013	0.002	-0.018**	-0.020
	(0.009)	(0.008)	(0.009)	(0.006)	(0.011)
StillEnrolled	-0.004	-0.016***	-0.005	-0.008*	-0.018*
	(0.005)	(0.005)	(0.006)	(0.003)	(0.008)
GradeProgression	-0.000	0.012	0.027	0.013	-0.012
	(0.015)	(0.013)	(0.017)	(0.009)	(0.022)
Years of Schooling	-0.146**	-0.074	0.076	-0.097**	-0.025
	(0.048)	(0.044)	(0.054)	(0.031)	(0.069)
N	161380	153011	79198	321140	72457
Primary school	-0.005	0.005	-0.012	-0.012	-0.005
	(0.026)	(0.022)	(0.025)	(0.017)	(0.026)
N	46759	46216	24832	95064	22771

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, gender, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of children in the family at the time of survey. Instrument is the a dummy variable indicating if the first-born is a female child. Low caste includes scheduled caste (SC) and scheduled tribe (ST). Middle caste is the other backward caste (OBC).

Table 8: 2SLS estimates of the effect of family size on education by mother's education and household wealth; 2+

	Mother's Education			Wealth Category		
	Illiterate	Less than Primary	Primary & above	Low	Middle	High
	(1)	(2)	(3)	(4)	(5)	
Lit	-0.044***	-0.026**	-0.003	-0.051***	-0.017*	-0.008
	(0.009)	(0.010)	(0.008)	(0.011)	(0.007)	(0.009)
EverSchool	-0.038***	-0.006	0.003	-0.039***	-0.009	0.003
	(0.008)	(0.007)	(0.005)	(0.010)	(0.005)	(0.005)
StillEnrolled	-0.016***	-0.017**	-0.004	-0.016**	-0.006	-0.009
	(0.005)	(0.006)	(0.003)	(0.005)	(0.004)	(0.005)
GradeProgression	-0.001	0.041*	-0.001	0.013	0.011	-0.013
	(0.012)	(0.018)	(0.017)	(0.014)	(0.012)	(0.019)
Years of schooling	-0.280***	0.076	0.069	-0.253***	-0.009	0.068
	(0.044)	(0.050)	(0.039)	(0.050)	(0.034)	(0.050)
N	227697	63815	102085	191211	154262	48124
Primary school	-0.053*	0.041	0.021	-0.007	-0.018	0.042**
	(0.024)	(0.025)	(0.015)	(0.037)	(0.016)	(0.015)
No of observations	70573	19818	27431	52844	49255	15718

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, gender, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of children in the family at the time of survey. Instrument is the a dummy variable indicating if the first-born is a female child.

Table 9: 2SLS estimates of the effect of family size on education by caste and residence; 3+

	Instrument: First two children are girl (GG)				
	Low Caste	Middle Caste	High Caste	Rural	Urban
	(1)	(2)	(3)	(4)	(5)
Lit	-0.020*	-0.020**	0.005	-0.017**	-0.008
	(0.009)	(0.008)	(0.011)	(0.005)	(0.013)
Ever School	-0.028***	-0.016*	-0.001	-0.018***	-0.023*
	(0.008)	(0.006)	(0.009)	(0.005)	(0.011)
Still Enrolled	-0.003	-0.006	0.000	-0.002	-0.014
	(0.004)	(0.004)	(0.006)	(0.003)	(0.007)
Grade Progression	-0.027*	-0.008	-0.014	-0.020**	-0.001
	(0.012)	(0.011)	(0.017)	(0.008)	(0.020)
Years of Schooling	-0.190***	-0.136***	-0.144**	-0.183***	-0.085
	(0.036)	(0.034)	(0.053)	(0.023)	(0.062)
N	161380	153011	79198	321140	72457
Primary school	-0.012	-0.018	-0.019	-0.023*	-0.017
	(0.016)	(0.015)	(0.019)	(0.010)	(0.023)
N	46759	46216	24832	95064	22771

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, gender, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of children in the family at the time of survey. Instrument is the a dummy variable indicating if the first-born is a female child. Low caste includes scheduled caste (SC) and scheduled tribe (ST). Middle caste is the other backward caste (OBC).

Table 10: 2SLS estimates of the effect of family size on education by mother's education and household wealth; 3+

Instrument: First two children are girl (GG)							
	Mother's Education				Wealth Category		
	Illiterate	Less than Primary	Primary & above		Low	Middle	High
	(1)	(2)	(3)		(4)	(5)	
Lit	-0.022***	-0.016	0.009	-0.023**	-0.010	0.008	
	(0.006)	(0.010)	(0.009)	(0.008)	(0.006)	(0.013)	
EverSchool	-0.025***	-0.015*	0.004	-0.027***	-0.014**	0.005	
	(0.006)	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)	
StillEnrolled	-0.007	-0.007	0.002	-0.004	-0.003	-0.009	
	(0.004)	(0.006)	(0.005)	(0.004)	(0.004)	(0.007)	
GradeProgression	-0.022*	0.011	-0.027	-0.019	-0.016	-0.003	
	(0.009)	(0.017)	(0.018)	(0.011)	(0.011)	(0.024)	
Years of schooling	-0.239***	-0.026	-0.034	-0.196***	-0.132***	0.046	
	(0.028)	(0.050)	(0.045)	(0.032)	(0.034)	(0.066)	
N	227697	63815	102085		191211	154262	48124
Primary school	-0.038**	0.029	-0.007	-0.013	-0.035**	0.028*	
	(0.013)	(0.018)	(0.014)	(0.017)	(0.013)	(0.014)	
No of observations	41950	10832	12265	30696	27174	7133	

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age square, gender, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of children in the family at the time of survey. Instrument is the a dummy variable indicating if the first-born is a female child.

The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey[#]

Sarojini Hirshleifer, *UCSD*

David McKenzie, *World Bank, BREAD, CEPR, and IZA*

Rita Almeida, *World Bank*

Cristobal Ridao-Cano, *World Bank*

Abstract

We use a randomized experiment to evaluate a large-scale active labor market policy: Turkey's vocational training programs for the unemployed. A detailed follow-up survey of a large sample with low attrition enables precise estimation of treatment impacts and their heterogeneity. The average impact of training on employment is positive, but close to zero and statistically insignificant, which is much lower than either program officials or applicants expected. Over the first year after training we do find training to have had statistically significant effects on the quality of employment, and that the positive impacts are stronger when training is offered by private providers. However, longer-term administrative data shows that after three years these effects have also dissipated.

Keywords: Vocational training; Active Labor Market Programs; Randomized Experiment; Private Provision.

JEL Codes: I28, J24, J68, O12, C93.

[#] We thank the Spanish Impact Evaluation Fund (SIEF), the Gender Action Plan, the World Bank's Research Support Budget, and ISKUR for co-funding this impact evaluation. We thank Ayca Donmez, Elcin Koc, and Levent Yener for their assistance on this project, audiences at a number of seminars for useful comments, and ISKUR for their collaboration. All views expressed in this paper are those of the authors alone and need not necessarily reflect those of the World Bank or ISKUR.

1. Introduction

After a decline in funding in the 1990s and early 2000s, vocational training programs began to return more prominently to the agendas of governments and international donor agencies in the mid-2000s (King and Palmer, 2010). An employer-identified unmet demand combined with constraints on the supply of skills in the working-age population has created a concern that low skill levels are impeding development in some countries (UNESCO, 2012; World Bank 2012a). The global economic recession that began in 2007 has also dramatically increased interest in policies that could be used to reduce unemployment (World Bank, 2012b). The persistence of labor market imbalances has led to the worry that unemployment is becoming more structural in nature, requiring an emphasis on skills training to help reduce skills mismatches (ILO, 2012). As a result, expanded training programs were the most common type of labor market policy implemented globally in response to the crisis (McKenzie and Robalino, 2010).

The key question is then: do such policies work in helping individuals who receive training to subsequently find jobs? A review of the U.S. literature by Heckman et al. (1999) found substantial heterogeneity in the impact of estimates of across studies and concluded that job training had at most modest positive impacts on adult earnings, with no impact or even negative impacts for youth.[1] Similarly, Kluve (2010) in reviewing evaluations of programs in Europe concludes that they also show, at best, modest positive effects, but that programs for youth are less likely to show positive impacts. In developing countries, there have been few rigorous evaluations of training programs. Job training may be more effective in developing countries, however, if a skills gap is especially likely to be the binding constraint to employment (Dar et al, 2004; World Bank 2012a). Recently, three randomized evaluations have been conducted of vocational training programs directed at disadvantaged youth in Colombia (Attanasio et al, 2011), the Dominican Republic (Card et al, 2011), and Malawi (Cho et al, 2013). The results in Malawi and the Dominican Republic are consistent with the earlier literature, with no impact on employment in either, and perhaps modest increases in income in the Dominican Republic. Somewhat more encouraging results are found in Colombia with young women offered training

[1] More recently, Schochet et al. (2008) find improvements in earnings for disadvantaged youth taking part in the Jobs Corps program.

having a 7 percent increase in employment and 20 percent increase in earnings, although men saw no change in these outcomes.[2]

This paper builds on and extends this literature by providing the first randomized experiment of a large-scale vocational training program for the general unemployed population (not just for disadvantaged youth) in a developing country. This is also the first paper from a developing country that is able trace longer-term impacts up to three years post-training, by complementing a follow-up survey with administrative data from the social security agency.. We employ an over-subscription design to evaluate the impact of the Turkish National Employment Agency's vocational training programs.[3] These programs average 336 hours over three months, are available for a wide range of subjects, and are offered by both private and public providers. These training services were provided to over 250,000 registered unemployed in 2011, hence we are evaluating a program operating at scale and not just a pilot. A large sample of 5,902 applicants randomly allocated to treatment and control within 130 separate courses, coupled with a detailed follow-up survey with only 6 percent attrition one year after training allows us to measure both the overall impact of training, and heterogeneity in training impacts along dimensions pre-specified in a pre-analysis plan. These features contrast with the small existing literature in developing countries, which has focused on pilot programs for youth, evaluated over relatively short time horizons, with higher rates of attrition, and no sources of administrative data on employment outcomes.

We estimate that being assigned to training had an overall effect on employment and earnings that was small in magnitude and not statistically significant: individuals assigned to treatment had a 2 percentage point higher likelihood of working at all, a 1.2 percentage point higher likelihood of working 20 hours or more per week, and earned 5.6 percent higher income. We find similar magnitude, but statistically significant, impacts on measures of the quality of employment: a 2.0 percentage point increase in formal employment, 8.6 percentage point increase in formal income, and an increase in occupational status. Our point estimates suggest the training impacts were largest for males aged above 25, even though this group was least likely to take a course conditional on being assigned to treatment. We cannot, however, reject

[2] In addition a small pilot study of 658 women offered tailoring and stitching training in India found a 5 percent increase in employment (Maitra and Mani, 2012), while two studies of youth in Uganda which offer vocational training in combination with a grant (Blattman et al, 2013) or in combination with life skills training (Bandiera et al, 2012) also found positive impacts on employment and/or earnings.

[3] This study is also to our knowledge the first randomized experiment of any social policy in Turkey.

equality of impacts by age and gender, nor do we find robust heterogeneity with respect to other individual characteristics. Consistent with these modest overall impacts, we do not find treated individuals to be in better mental health, to be any more likely to expect to be working in 2 years time, or to expect a higher future subjective well-being than individuals who are not trained. An expectations elicitation exercise reveals these impacts to be substantially smaller than anticipated by either program applicants or Employment Agency staff. Using administrative data to examine longer-term impacts, we confirm a small, but positive significant impact on formal employment over a one year horizon, which dissipates over time so that there is no significant impact on formal employment three years after training.

We then examine heterogeneity of impacts along a pre-specified causal chain of training impact in an attempt to understand why the impacts were lower than expected, and under what circumstances they are higher. We test which course characteristics associated with training quality matter, and we that find attendance rates for courses are relatively high, and there is little evidence of heterogeneity by course length, teacher quality measures, participant assessments of what the course teaches, or by the unemployment rates in the labor markets in which courses are taught. Instead, we find training to have stronger impacts when offered by private providers, with this heterogeneity being robust to adjustments for multiple hypothesis testing. Controlling for observable differences in other course characteristics, local labor market conditions and the applicants for different courses does not change this conclusion. This finding that impacts are higher with private provision is consistent with some non-experimental work (e.g. Jespersen et al, 2008) but to our knowledge this is the first time it has been found in an experimental setting. However, longer-term administrative data show that the private provider effect is no longer significant in the medium term. As a result, cost-benefit analysis suggests that even privately-run courses struggle to provide positive returns.

The remainder of the paper is structured as follows: Section 2 discusses the context and Turkey's vocational training courses, Section 3 the experimental design, and Section 4 our data collection and estimation methodology. The main results are presented in Section 5, while Section 6 traces out a causal chain from training to employment. Section 7 examines in more detail the differences between private and public courses, and Section 8 discusses cost-effectiveness. Section 9 concludes.

2. Context and Turkey's Vocational Training Courses

Turkey is a middle income country with a population of almost 75 million, and per capita income of US$10,400 ($17,300 in PPP terms). Urbanization, income and unemployment vary significantly at the province level. Nationwide, the population is 70% urban. In 2010, the employment rate for 15-64 year olds in Turkey was 46.3 percent compared to an E.U. average of 64.1 percent and a rate of 66.7 percent in the U.S.[4] This low employment rate is driven partly by a female employment rate of only 23 percent, but also reflects relatively high unemployment. The unemployment rate was 12.5 percent in 2009, and as is common in much of the world, was much higher for youth than for older workers (16.8 percent for youth under 25 in 2011, versus 7.2 percent for 25-64 year olds].[5]

The Turkish National Employment Agency (ISKUR) provides services for individuals who register as unemployed through 109 offices in 81 provinces. Training programs are the primary active labor market program provided by ISKUR.[6] The Turkish Government dramatically increased access to ISKUR-supported training programs from 32,000 individuals trained in 2008 to 214,000 in 2010 and 250,000 in 2011. This expansion began as a means to mitigate the impact of labor market reforms and continued as a response to a spike in unemployment that coincided with the 2009 global recession.[7]

The majority of training courses offered are general vocational training courses covering a wide range of vocations[8]. These are contracted to a mix of public and licensed private providers. Courses are announced in ISKUR offices, on the ISKUR website, and by text messages. The courses are provided free to the trainees, and the trainees also receive a small stipend of 15 Turkish Lira (US$10 in 2010) per day during the course period (which averages three months). Given excess demand for courses and a desire to get the unemployed into jobs, individuals are

[4] Source: Eurostat http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Employment_statistics [accessed December 14, 2012].

[5] Source: Eurostat http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Unemployment_statistics [accessed December 14, 2012].

[6] ISKUR provides information about job openings to the unemployed, but does not have the intensive job search support and counseling that is often provided to the unemployed in rich countries.

[7] Note that Turkey experienced a contraction in 2009, but the economy then recovered quickly, with 9 percent growth in 2010 and 8.5 percent in 2011, and the unemployment rate dropped back to the average pre-crisis level by February 2011.

[8] ISKUR also offers training programs designed to provide convicts and ex-convicts with skills to enter the labor market after release; courses for starting a business; a small public works program; and on-the-job training programs. General vocational training accounted for 61 percent of the total number of participants in all of ISKUR's active labor market programs in 2009.

only allowed to take one ISKUR-supported course in a five-year period. To be eligible to participate in the course, individuals must be at least 15 years in age, have at least primary education, and meet other skill pre-requisites which depend on the course they wish to participate in (for example, software courses may require some pre-existing IT knowledge or skills).

3. Experimental Design

During this period of rapid expansion of provision of vocational training services, the Turkish Ministry of Labor asked the World Bank for assistance in evaluating the impact of these courses. Excess demand among the unemployed for many of the courses offered by ISKUR provided the possibility for an oversubscription design. Moreover, it enabled evaluation of courses being offered at scale, rather than on a pilot basis.

3.1 Selection of Provinces

Our desire was to choose provinces in order to ensure a broad geographic distribution and range of labor market conditions. Selection of provinces to conduct the evaluation in began with a list of the 39 provinces which had had at least two significantly oversubscribed training courses in 2009. These provinces were first stratified by whether they had an unemployment rate above or below the median of 10 percent in 2009. Ten provinces were then randomly selected from each strata with probability proportional to the percentage of individuals trained in 2009. Three additional provinces (Antalya, Gaziantep, and Diyarbakir) were included in the sample at the request of ISKUR because of their importance in representing varying labor market conditions across Turkey. As a result, 23 provinces were selected for inclusion in the evaluation (Figure 1).

3.2 Selection of Evaluation Courses

Power calculations gave a target sample size of 5,700 individuals. This target was divided amongst the 23 provinces in proportion to the number of trainees in these provinces in the previous year. Thus Istanbul accounts for 21.8 percent of the sample, Kocaeli, Ankara and Hatay collectively another 28 percent, and the remaining half of the sample is split among the other 19 provinces.

The evaluation team worked with regional ISKUR offices to determine the actual courses from within each province to be included in the evaluation. The key criteria used to decide which courses to include in the evaluation were i) the likelihood of the course being oversubscribed

(which ensures the most popular types of training, for which there would be demand for further scale-up, are included); ii) inclusion of a diversity of types of training providers to enable comparison of private and public course provision; and iii) course starting and ending dates. The evaluation includes courses that started between October and December 2010 and finished by May 2011 (75 percent had finished by the end of February 2011). The timing of the evaluation was determined by the fact that it tends to be a time of year when people in Turkey are more likely to seek training through ISKUR.

This resulted in a set of 130 evaluation courses spread throughout Turkey, of which 39 were offered by private providers and the remainder were mainly government-operated. The single most common course was computerized accounting, which 24 percent of trainees applied for. Twenty one percent of trainees were in service courses (babysitter, cashier, waiter, caring for the elderly), 15.4 percent were craftsman or machine operators (welder, natural gas fitter, plumber, mechanic), 14.7 percent were in technical courses (computer technicians, computer-aided design, electrical engineering), and 12.2 percent were in professional courses (web designers, computer programmers, IT support specialists). The average course size was 28 trainees, and the average course length was three months (typically around 6 hours per day), both with significant variation.[9] Three months is the same as the average length of the classroom components in both the Colombian and Dominican Republic training evaluations (Attanasio et al, 2011; Card et al, 2011), although those programs also supplemented this with two to three month internships to provide further on-the-job training, and is also as long as the apprenticeship program to teach vocational skills in the Malawi study (Cho et al, 2013).

3.3 Assignment of Individuals to Treatment and Control within Courses

Courses were advertised and potential trainees applied to them following standard procedures. Applications were then screened to ensure they met the eligibility criteria of ISKUR and the course provider. Training providers were then asked to select a list of potential trainees that was at least 2.2 times capacity. Typically this involved short interviews with eligible applicants. Courses in which between 1.8 and 2.2 times the course capacity were deemed suitable candidates were also included, although in those cases less than the full class size was allocated to

[9] The maximum course length in our sample is 128 days, with 90 percent of course participants in courses of 90 days or fewer.

treatment. These individuals' applications details were then submitted into ISKUR's Management Information System (MIS).

The MIS system then stratified applicants for each course by gender and whether or not they were aged less than 25. Within these strata, the MIS then randomly allocated trainees at the individual level into one of three groups: a treatment group who were selected for training, a control group who were not, and a waitlisted group who the training provider could select into the training if there were drop-outs. Since training providers are paid on the basis of number actually trained, if individuals assigned to treatment drop out of training, providers look to quickly fill in the empty spots. In Card et al. (2011)'s evaluation in the Dominican Republic, this led to one-third of the control group being offered treatment, with this selection typically non-random. The inclusion of a waitlist was done to prevent this from occurring. Thus if a course had capacity for 50 trainees, and 120 were deemed eligible, 50 would be randomly assigned to treatment, 50 to control, and 20 to a waitlist.[10] If individuals from the treatment group dropped out before one-tenth of the training course had been completed, providers could draw replacements as they liked from the waitlist. If they exhausted the first waitlist, then they drew from a second waitlist of individuals who had just missed the cut of being in the top group of applicants. At no time were trainees or prospective trainees informed of the evaluation. We do not use the waitlist in our study, since their selection into training is non-random.

The final evaluation sample consists of 5,902 applicants, of which 3,001 were allocated to treatment and 2,901 to control. There are 173 individuals who applied to more than one course. These individuals were still randomly selected into treatment or control, but have a higher probability of selection since they participate in more than one course lottery. Our estimation strategy accounts for this.

3.4 Compliance with Treatment

As is common with many training programs, not all those accepted into the course took up the training. Twenty-three percent of the individuals assigned to treatment chose either to not take training or had dropped by the second day of the course, despite it being typically only two to three weeks between the interviews for a course and its start date. There was relatively little

[10] If between 1.8 and 2.2 times the course capacity applied, then proportionally smaller groups were chosen. For example, if 100 people applied for a course with capacity 50, then 50 individuals would be allocated to the training (of which 40 would be designated the treatment group for surveying and measurement purposes), 40 would be allocated to control, and 10 to the waitlist.

further dropout during the course; 72% of the treatment group completed training and 69% of the treatment group received a certification of course completion in an evaluation course. Compliance does vary substantially by sub-group, with 72% of women and 62% of men receiving certification; people under 25 were also more likely to complete training. Another 3% of the treatment group as well as 3% of the control group received certification in an ISKUR course that was not included in the sample of evaluation courses during the period that evaluation courses took place.[11]

4. Data Collection, Baseline Comparisons, and Estimation Methodology

The MIS system contained basic information about the course and the sex, age, and education level of the applicant. The main data for evaluation then come from surveys administered to the applicants, with some supplementary data from a survey of training providers, and some longer-term information on formal employment from the social security system.

4.1. Surveys

A baseline survey and follow-up survey were both conducted through in-person interviews by a third-party professional survey firm (Frekans) that was not affiliated by ISKUR and which was selected by the evaluation team. The baseline survey took place on a rolling basis between 13 September, 2010 and 31 January, 2011. The goal was to conduct the surveys before courses began, but given the short window of time between selection of applicants and the start of the course, in practice only one-third of those surveyed were surveyed before the start of the course, while 79 percent of those interviewed were interviewed with 11 days of the start of the course. Applicants were told that the purpose of the survey was to help improve the services offered by ISKUR, and that their participation in the survey had no impact on being accepted into any training course, nor would their data be shared with ISKUR at the individual level. The overall baseline response rate was 90 percent.

The follow-up survey took place between December 27, 2011 and March 5, 2012, which corresponds to a period approximately one year after the end of training. It collected data on employment outcomes, as well as individual and household well-being. The response rate was 94

[11] An additional 7.3% of the control group and 2.7% of the treatment group took a non-ISKUR course at any time from the beginning of the study period up to the time of the baseline survey. The majority of these courses were funded by other government entities. Our power calculations accounted the possibility that at least 15% might ultimately take up training.

percent, including 472 individuals who had not been able to be interviewed at baseline. In total 5,057 individuals were interviewed at both baseline and follow-up.

The attrition rate of 6 percent at follow-up compares favorably with the attrition rates in other evaluations of vocational training in developing countries (18.5 percent in Attanasio et al, 2011; 23 percent in Maitra and Mani, 2012; 38 percent in Card et al, 2011, and 46 percent in Cho et al, 2013). Table 1 examines whether attrition from either the baseline or follow-up surveys is related to treatment status, both for the full sample, and then by the four gender by age group stratum. To do this, we estimate the following regression:

$$Attrition_i = \alpha + \beta AssignedtoTraining + \sum_{s=1}^{457} \theta_s \delta_{i,s} + \varepsilon_i \quad (1)$$

Where i denotes individuals, s denotes a course*gender*age group lottery, and $\delta_{i,s}$ is a dummy variable indicating whether individual i applied for lottery s.[12] This controls for randomization strata (Bruhn and McKenzie, 2009) as well as controlling for individuals who applied for more than one course (Abdulkadiroglu et al, 2011). We are then interested in β, the impact of being assigned to receive vocational training, on attrition.

Table 1 reports the results. Attrition at baseline is 2.7 percentage points lower for the treatment group than the control, while attrition at follow-up is 1.4 percentage points lower. This differential attrition comes from relatively higher attrition from males in the control group than for females or for treated males, while there is no significant differential attrition for females. Whilst statistically significant, this difference in attrition rates by treatment status is small in absolute terms. We examine the sensitivity of our results to differential attrition by employing Lee (2009)'s bounding approach. In addition, attrition is near zero in our administrative data (only 6 individuals could not be matched). This gives us several outcome measures which are not subject to concerns about attrition. Furthermore, using this administrative data, we are unable to reject that attrition rates in the follow-up survey are unrelated to the likelihood of having found a formal job by the time of this survey (p=0.807).

4.2 Administrative Data on Employment from Social Security Records

Our sample of applicants was linked by ISKUR to official data on worker payments filed in the social security system. Due to updates and delays in this system, it took us two years to receive this data from when it was requested, and we were limited in the information that could

[12] There are 457 strata, reflecting the non-empty cells from 130 courses*2 genders*2 age groups.

be extracted. First, we received formal employment status, and earnings reported for social security, for the month of August 2013. This corresponds to a time period of approximately one and a half years after our follow-up survey, and two and a half years after the end of training. Secondly, in December 2013, we were given data on the first date that a worker was registered in the social security system after the end date of the course which he or she had applied for, as well as the date that they left this job, if this job ended. This data effectively covers the period from January 2011 up to the end of November 2013. It tells us month by month whether an individual was ever formally employed post-training, but since it does not report on second and subsequent formal jobs, not current formal employment status. We use this data to measure whether an individual had been formally employed by January 2012 (the mid-point of the follow-up survey), as well as to examine the trajectory of entry into formal employment.

Although we we cannot fully capture the longer-term impactof training on total employment, our follow up survey indicates that, conditional on being employed at 20 hours a week, 82% the sample is formally employed. In addition, to the extent that training has any impact in measures from the follow up survey, it is concentrated in employment quality measures, such as formal employment

4.3 Balance and Baseline Characteristics

Random assignment to treatment and control occurred at the individual level and was done by computer using code written by the evaluation team. Consistent with this, the first few columns of Table 2 show balance on course and demographic characteristics using the administrative data available for the full sample. The remaining columns then examine balance for the sample interviewed at baseline, and for the sample interviewed at follow-up. Given that some baseline interviews took place after the course had started, we focus on comparing either time invariant or slow-moving individual characteristics collected during the baseline survey. The differences between treatment and control are small and magnitude, and the only significant difference at the 10 percent level is that the treated group surveyed at baseline is marginally less likely to have worked before. Given the number of variables tested, we view this as the result of chance, and that randomization has succeeded in providing comparable treatment and control groups. Applicants for the evaluation courses are, in general, relatively young and well-educated. The average applicant is 27 years old, and approximately 73 percent of them have completed high

school. Sixty-one percent of the evaluation sample is female. The majority have worked previously, with 63 percent having worked before, and an average work experience of more than three years. Nonetheless, 37 percent have never worked before, and a further 20 percent have worked for less than one year. Baseline responses to current employment are consistent with these individuals being unemployed: only 12 percent have worked even one hour in the previous month, and only 2 percent have worked at least 20 hours per week in the past month.

A comparison of the applicants to a sample of all urban unemployed using the 2009 Labor Force survey reveals that applicants are younger, more likely to be female, have less work experience, and are better educated than the average unemployed individual. Only 30 percent of the unemployed are female (reflecting low labor force participation rates among women), compared to 61 percent of applicants; only 31 percent of the unemployed are youth, compared to 45 percent of applicants; and only 42 percent have completed high school, compared to 73 percent of applicants. These differences reflect both differences in which types of unemployed individuals are more likely to apply for training courses, as well as the supply of courses, since many courses are designed for people with at least medium levels of education. Our results are therefore informative as to the effect of training for the types of unemployed who apply for vocational training courses being offered, but need not necessarily reflect the returns to vocational training for the broader population of the unemployed.

4.4 Estimation Methodology

We can measure the intent-to-treat (ITT) effect of vocational training on a particular outcome of interest by estimating the following equation, analogous to equation (1) :

$$Outcome_i = \alpha + \beta AssignedtoTraining + \sum_{s=1}^{457} \theta_s \delta_{i,s} + \varepsilon_i \qquad (2)$$

We do not control for baseline levels of outcomes since they are close to zero and provide little variation. β is then the average effect of being selected for a vocational training course on this outcome.

We can also estimate the impact of actually completing training by replacing *AssignedtoTraining* with *CompletedTraining* in (2), and instrumenting this with treatment assignment. Under the assumption that assignment to training has no impact on outcomes for those who do not complete the course, and that there are no individuals who would take courses only if assigned to the control group, this yields the local average treatment effect (LATE). This

is the impact of completing training for an individual who takes up training when they are selected in the course lottery, and does not take it up otherwise. A concern with this estimation is the possibility that simply being selected for a course may affect employment outcomes, even if individuals do not take the course or drop out after only a few days. For example, individuals (or employers) may take selection for the course as a signal of quality, which may give selected applicants more confidence approaching employers or employers a spur to hire these individuals. Showing up and finding out that you don't like the course may cause reluctant job-seekers to try harder to look for work. Due to these concerns, we focus on the ITT estimates for most of our analysis, and just report the LATE estimates for our overall employment outcomes.

The primary outcomes of interest relate to employment. We consider a variety of employment measures which aim to measure whether individuals are employed at all, as well as how much they are working, how much they earn from this work, and the quality and formality of this employment. Appendix A explains how our key variables were constructed. We pre-specified these outcomes and how they would be measured in a pre-analysis plan (Casey et al, 2012) that was archived on February 6, 2012, before any follow-up survey data was received.[13] To control further for multiple hypothesis testing among employment outcomes, we follow Kling et al. (2007) in estimating a mean treatment effect on an aggregated employment index. This first transforms each employment outcome by subtracting the mean and dividing by the standard deviation, and then takes an average across outcomes.

In addition to estimating the overall impact of training, we are interested in exploring the heterogeneity of impacts to help understand whether certain types of courses offer larger impacts, or certain types of people benefit more from training. Our pre-analysis plan specified dimensions of heterogeneity of interest. To estimate heterogeneity with respect to characteristic X, we estimate

$$Outcome_i = \alpha + \beta AssignedtoTraining + \gamma AssignedtoTraining * X + \rho X + \sum_{s=1}^{457} \theta_s \delta_{i,s} + \varepsilon_i \qquad (3)$$

Recently Fink et al. (2012) have criticized randomized experiments looking for heterogeneity in treatment effects for not controlling adequately for multiple hypothesis testing. They recommend the use of the Benjamini and Hochberg (1995) approach which holds constant the

[13] The pre-analysis plan was archived and time-stamped by the J-PAL Hypothesis Registry, and is available at http://www.povertyactionlab.org/sites/default/files/documents/ISKURIE_AnalysisPlan_v4.pdf

false discovery rate (the expected proportion of falsely rejected null hypotheses). We use this approach to examine which dimensions of heterogeneity are robust to this concern.

5. Impacts on Employment and Well-being

We begin by looking at the overall impacts on employment for the pooled sample, and then look separately by the four age*gender strata, and for heterogeneity with respect to pre-specified human capital variables. We then examine impacts on measures of current and expected future well-being.

5.1 Overall Impacts on Employment

Table 3 reports the results of estimating equation (2) for different employment outcomes measured in our follow-up survey data. The first two columns show small and insignificant positive impacts of training on the likelihood of working at all, and of working at least 20 hours or more a week. The ITT for working at all is for a 2 percentage point impact, with a 95 percent confidence interval of [-0.5, +4.4] percentage points. This is thus a relatively precise zero or small effect. Lee bounds to adjust for differential attrition are also fairly narrow, consistent with the amount of differential attrition being small. The LATE estimates are for a 2.9 percentage point increase in being employed at all, and a 1.9 percentage point increase in the likelihood of working at least 20 hours per week.

We also find small and statistically insignificant impacts on weekly hours worked, income from work, and on a transform of income from work which is less sensitive to outliers. In contrast, we find statistically significant (at the 10 percent level) impacts on several measures of job quality: the socioeconomic occupational status of the job, being formally employed, and the income earned from formal jobs. For example, we find a 2 percentage point increase from being assigned to training in the likelihood of being formally employed, with the LATE impact being 3 percentage points. Given that 29 percent of the control group is formally employed at follow-up, this LATE estimate is equivalent to a 10 percent increase in formal sector employment.

All of these different employment outcomes show modest positive impacts.[14] Averaging them together to account for multiple testing and examine the mean impact finds a positive overall impact which is significant at the 10 percent level. However, the size of the impact is

[14] The modest size of the treatment effect cannot be explained by trainees being more likely to pursue additional education and training. The impact of treatment on the probability of being in education or training at the time of the follow-up survey is very small, negative and marginally significant (-.014, p-value = .095)

small, with the ITT showing only a 0.04 standard deviation improvement in employment outcomes as a result of being selected for training.

Table 4 uses the social security data to measure impacts on formal employment. Consistent with our survey data, the first column shows a statistically significant 2.6 percentage point impact on the likelihood of having found a formal job by the time of the follow-up survey. However, the next three columns show that this effect disappears over time. There is no significant effect on formal employment status or formal income by August 2013, and no impact on whether they have ever found a formal job by the end of November 2013. Figure 2 examines the trajectory of impacts in more detail, plotting the proportion of treatment and control who have ever found a formal job month by month. We see that the treatment group is more likely to have found a formal job than the control group by May 2012, and that this gap lasts for about a year, including the time of the follow-up survey, before closing again.[15] By November 2013, 66 percent of both groups have found a formal job at some stage during the post-training period. However, there appears to be a lot of churn in employment, with the last column of Table 4 showing that 50 percent of both groups have also left a formal job at some point during this period. As a result, there is no lasting impact of training on formal employment.

5.2 Heterogeneity of Employment Impacts by Age and Gender

Existing experimental evaluations of vocational training programs in developing countries have focused just on youth, and Attanasio et al. (2011) find different impacts by gender, with young women benefiting more from their training in Colombia than young men. Recall that our randomization stratified by gender*age group within each course. We estimate equation (3) with X as the vector of 4 gender *age group strata and report the results in Table 5. A test of equality of treatment effects across the four subgroups enables us to determine whether there is significant heterogeneity in employment effects by age or gender. Lee bounds control for the differential attrition, which is largely only an issue for males.

The key result from Table 5 is that we cannot reject the null hypothesis of equality of employment treatment effects across the four age*gender groups. This is despite the four groups

[15] The control group was about 5 percentage points more likely than the control group to take a non-ISKUR training course. If those individuals entered the labor force later, it could at least partially explain the lag in the control entering the labor force. Another possible explanation is that the impact of training interacted with changing labor market conditions.

having quite different employment experiences over the year since courses ended: 62.5 percent of men aged 25 and older have worked in the past month, compared to 48.9 percent of younger men, 40.3 percent of younger women, and only 29.5 percent of women aged 25 and older. The lack of a strong employment effect of training is thus not because the labor market is so weak that no one can find jobs regardless of whether or not they are trained.

If we were to look at the four subgroups separately, only males aged 25 and older show significant treatment impacts on some employment outcomes, which appear robust to Lee bounding for differential attrition. Men in this age group assigned to training are 6.9 percentage points more likely to be working, are working 2.9 hours more per week on average, and are in a higher average occupational status (although this captures both the impacts at the extensive margin (working or not) and intensive margin (jobs taken up conditional on working). The Benjamini and Hochberg (1995) sequential adjustment approach would still show the impact on working at all and on occupational status to be significant if the false discovery rate across the four age*gender subgroups is held constant at 10 percent, but not significant if a 5 percent FDR is maintained. Thus there is evidence to support an impact for males aged over 25, but we also cannot reject that this impact is not different from the impacts for the other three age*gender subgroups.

5.3 Heterogeneity of Impacts with Respect to Other Individual Characteristics

Next we examine further whether the small overall average treatment effect is masking large heterogeneity of responses across individuals with different types of human capital. If human capital and training are complements, then we would expect larger treatment effects for individuals with higher beginning levels of human capital. In contrast, if training substitutes or compensates for other forms of human capital which individuals are lacking, then we should expect larger treatment effects for individuals with lower initial human capital levels.

We consider a wide range of pre-specified measures of human capital, including education, measures of cognitive ability (Raven test and numeracy), empowerment, personality characteristics (work centrality and tenacity), and long-term unemployment (which may indicate deteriorated skills). In addition we examine heterogeneity with respect to an individual's expectations of how much they will benefit from the course (see section 5.5), and to the presence of young children in the household (which may change an individual's cost of working).

Although these measures do help predict levels of employment, Table 6 shows there is very limited heterogeneity of treatment impacts with respect to any of them. Individuals who have previously taken a training course before, and those who are the main decision-makers about work, have larger impacts of the training according to our survey outcomes. However, the significance of neither variable survives corrections for multiple hypothesis testing, nor do we see significant impacts of these variables on the administrative measures of employment. The most statistically significant coefficient is that on tenacity, which has a negative impact on the training effect on formal employment in August 2013. This could reflect that tenacious individuals will find employment anyway, so that training substitutes for a lack of this characteristic. However, the p-value of 0.010 is not small enough to survive a correction for testing heterogeneity across so many different measures.

5.4 Impacts on Well-being

Our survey measured impacts approximately one year after the completion of training. This is similar to the timing in other experimental vocational training evaluations in developing countries, and should allow long enough for most individuals to use their training in finding a job. Nevertheless, the concern remains that perhaps some of the impacts will take time to manifest. If that is the case, we should expect to see that training leads individuals to think their future job prospects and future household well-being will be better, even if there is relatively little change in current well-being. We test this in Table 7.

While 42 percent of the control group is currently working at all at the time of the follow-up survey, column 1 shows that 54.1 percent expect to be employed in two years time. However, there is no significant impact of treatment on this expectation. Column 2 shows that training does not improve mental health. Column 3 shows a marginally significant increase in current subjective well-being, although the impact is very small in magnitude: 0.07 steps on a 10 step-ladder, equivalent to 0.04 standard deviations. While control applicants do believe that their households will be on a higher step in five years than they are today, treatment has no additional impact on this. Treated individuals also do not have significantly higher household income, but they do have significantly higher durable asset levels. Again the magnitude of the impact is small, equivalent to 0.06 standard deviations.

Overall these results are therefore consistent with at best modest increases in employment translating into modest increases in current household wealth and subjective well-being, with no impacts on expectations about the future.[16] This accords with our finding from the administrative data that any (formal) employment impacts are not long-lasting. .

5.5 How do these results compare to the expectations of participants and policymakers?

To address the issue of whether these impacts are in line with the expectations of both the individuals applying for ISKUR training courses and the policymakers in charge of these courses, we conducted an expectations elicitation exercise (Groh et al, 2012). Our baseline survey elicited subjective probabilistic expectations (Delavande et al, 2010) by asking participants what they thought was the percent chance they would be employed in one year if they were selected for ISKUR training, and if they were not. A sample of 51 ISKUR employees from across the different provinces were also asked a range question about the likelihood the control group would be employed, along with a question eliciting the percent difference in employment rates they would expect from training.

Table 8 reports the results. The first two columns show that on average individuals were reasonably close in terms of their expectations of what employment levels would be like in the absence of training: the mean response was for a 31 percent chance of being employed, and in practice 36 percent of the control group is employed. Males aged 25 and over seem to underestimate their higher likelihood of being employed than the other groups, while older females do have lower expectations of being employed to match their lower realized levels.

In contrast to these relatively accurate expectations of the likelihoods of being employed without training, individuals dramatically overestimate the benefits from training: our LATE estimate is for a 2 percentage point increase in employment as a result of training, whereas the mean expected increase among those assigned to the treatment group is for a 32.4 percentage point increase.[17] This overestimation occurs for all gender and age groups. Moreover, ISKUR

[16] Our pre-analysis plan also hypothesized that training may further impact on social outcomes such as whether individuals head their own household, their decision-making power, and their attitudes towards women's role in society. We find no significant effects on any of these measures, which is consistent with the lack of strong labor market impacts (results available upon request).

[17] We report these expectations for the treatment group, to offset any concerns that individuals in the control group who had found out they were not selected for training by the time of the interview might understate their expectations of the value of training. In practice, however, the control and treatment group have very similar expectations, suggesting this is not a factor: the mean treatment gain expected by the control group is 31.9 percentage points, which is similar in magnitude to the 32.4 percentage point gain expected by the treatment group.

staff also overestimated the gain from training, with a mean expected treatment gain of 24.3 percentage points. The impact of training is thus much less than both the staff at ISKUR and the training applicants anticipated.

6. Why does training have such limited effects, and do certain types of training work better?

The impact of training is less than expected by either training participants or the labor ministry staff, although the impacts on employment are not that different from those found in other evaluations. In order for training to be effective, our pre-analysis plan set out four intermediate steps in a causal chain through which we might expect to see selection into a course influencing employment outcomes. We examine each in turn, which also helps understand whether certain types of training have more impact.[18]

6.1 Individuals selected for courses must show up and complete training

The first step in our causal chain is that individuals selected for training classes must actually show up and attend these classes. As noted above, 77 percent of those selected for treatment attended the course beyond the second day, and 72 percent completed it. The LATE estimates show impacts which adjust for attendance. However, we might also think that lower attendance and completion rates are an indicator of courses that participants find to be of lower quality or expect to be less useful. We therefore estimate equation (3) by interacting treatment with the percent of individuals assigned to a course who attended the course or who completed it. Since this characteristic varies only at the course level, we cluster standard errors by course (and also do this for all subsequent tests of interactions of treatment with course characteristics).

The first two rows of Table 9 show these treatment interactions. The interactions are positive, as hypothesized, suggesting that treatment effects are larger in courses with higher attendance rates. However, the results are almost all not statistically significant, and the magnitude of the impacts is not large: a one standard deviation increase in course attendance rates is associated with a 1 percentage point increase in employment. The administrative data shows a significant longer-term impact at the 10 percent level of the course having higher attendance, but the effect

[18] We specified that we would look at these effects for two survey outcomes: being employed for 20+ hours in a week, and the aggregate employment index. Given that we find impacts vary over time, we also examine this heterogeneity for two measures of formal employment from our administrative data: ever having a formal job between course end and January 2012 (around the follow-up survey), and being formally employed as of August 2013.

is still small in magnitude (a 1 s.d. increase is associated with a 2 percentage point increase in formal employment in August 2013), and is no longer significant after correcting for multiple hypothesis testing.

6.2 Higher quality courses should have more impact

A second concern is whether courses averaging three months in length are long enough to teach the skills necessary to improve labor market performance. The median course length is 320 hours which does suggest enough hours to enable learning to take place. To investigate whether longer courses have bigger impacts, we interact whether or not the course is above median length with treatment. The third row of Table 9 shows that longer courses in fact have less impact on employment than shorter courses amongst the sample of courses offered here. One possible explanation for this is that individuals reduce job search whilst taking part in training, so individuals in longer training courses have had less time to look for jobs (Rosholm and Skipper, 2009). The point estimates are significant at the 10 percent level, but are not significant after controlling for multiple hypothesis testing. They are certainly not consistent with the view that the limited impacts are due to the courses being too short, and the long term results from administrative data also do not show better impact from longer courses.

As proxies for the qualities of the teachers, we examine in the next two rows of Table 9 the treatment interactions with whether the average experience of teachers in the course is above 12 months (the median), and with the percentage of teachers for the course who are tertiary educated. We find very small and statistically insignificant impacts of either measure. This is consistent with much of the work in the education literature, and likely reflects the finding there that education and experience explain little of the actual variation in teacher effectiveness (Hanushek and Rivkin, 2006). It does suggest that the reason for limited impacts is not that the courses are taught by staff that are not experienced or educated enough.

Previous non-experimental literature in developed countries has found some evidence to suggest that the impacts of training are higher for training offered by private providers (Jespersen et al, 2008). Possible reasons are that private training providers are more responsive to private sector employer demand and/or they potentially face more competition and thus must increase quality in response. Rows six and seven of Table 9 examine the treatment interaction with facing two or more competitors, and with being a private provider. We see a marginally significant positive impact of facing some competition on our aggregate employment index. Private

provision has stronger impacts, with a positive impact on employment and on having had a formal job by the time of survey which are significant at the 10 percent level, and a positive impact on the aggregate employment index which is significant at the 1 percent level. The latter survives adjustments for multiple hypothesis testing. Training results in a 4 to 6 percentage point higher increase in employment after 1 year in private courses than in public courses. This gap narrows to 2.5 percentage points after 3 years, and is no longer statistically significant.

6.3 Skill Acquisition, Signaling, or Job Matching

A third step in the causal chain is for individuals who take courses of sufficient quality to be able to use what they have learned in the course to find jobs. There are three main channels through which vocational education may help in this respect. First, it might increase human capital through teaching new technical skills. Second, it may act to certify skills that individuals already have and act as a signaling mechanism to employers. Third, it may teach individuals new strategies for finding jobs in a certain profession, or better alert them to job opportunities, thereby improving job matching.

To examine the extent to which courses are playing each of these roles, and to which treatment effects vary with them, our follow-up survey asked course participants whether they thought the course had done each of these three things. We interact the percentage of course participants who think that the course taught new technical skills, certified existing skills, taught new strategies for finding jobs, or made them more aware of job opportunities with treatment and show the results in Table 9. We see that on average 84 percent of participants thought the courses certified skills they already had, and 80 percent thought they taught new skills, while 60 percent thought the course helped with job finding strategies and 45 percent with making them more aware of job opportunities. However, we find the point estimates typically to be negative and not statistically significant. Thinking the course made them aware of new jobs has a significant negative interaction for the outcome of having had a formal job at the time of the follow-up survey, but this is not significant after correcting for multiple testing. Thus the heterogeneity in how participants perceive these factors across courses does not seem to drive heterogeneity in treatment outcomes.

6.4 Training impacts and unemployment rates

The last step specified in the causal chain is for individuals who receive training to find jobs that they would not otherwise get. This depends on the labor market they face. On one hand,

when unemployment rates are low there should be more job opportunities available, making it easier for people to use their training to find jobs. But in such cases firms facing labor shortages might hire workers regardless of whether or not they have been trained. In contrast, when unemployment rates are high, if employers are not hiring then having new skills may not help the unemployed find jobs, but on the other hand, employers may be choosier and so training may offer workers a way to distinguish themselves from other workers competing for the same jobs.

As a result, theoretically it is ambiguous whether we should expect training to have more or less impact in situations of higher or lower unemployment. Consistent with this, there are mixed results in the existing (non-experimental) literature. Lechner and Wunsch (2009) find German training programs to be more effective when carried out during times when unemployment rates are higher and Kluve (2010) finds in his meta-analysis that program effects tend to be higher when unemployment rates are higher. But using Norwegian data, Raaum et al. (2003) find training impacts to be positively correlated with post-training employment rates in local labor markets.

The training programs in our study took place throughout 23 provinces with a variety of unemployment conditions – at the time of training, unemployment rates ranged from under 5 percent to over 20 percent, with a median of approximately 14 percent. The third to last row of Table 9 interacts the treatment with an indicator for above median provincial unemployment. The point estimates for our survey outcomes are positive, suggesting larger treatment impacts when labor markets are tighter, but not statistically significant. The longer-term impact is statistically significant, suggesting training to increase employment by 4.9 percentage points more when done in high unemployment regions compared to low unemployment regions, although this result does not survive a multiple hypothesis test correction.

Finally we examine heterogeneity with respect to the type of course. There is some evidence that accounting courses, the most common course type, have larger impacts on short-term employment outcomes than other types of courses. This impact is however not statistically significant in terms of formal employment, or long-term formal employment.

7. What distinguishes private courses from public courses?

The main source of impact heterogeneity in course characteristics is thus whether the course is operated by a private or public provider. Since individuals were only randomly assigned within courses, and not across courses, this might capture differences in the types of individuals

who apply for the different courses rather than of private provision per se. Moreover, even if we could randomly assign individuals to which course they took (which would be neither feasible nor desirable since it would mean assigning people to courses they did not meet the prerequisites for or had no interest in), private course providers still self-select into which courses they choose to offer, so we still could not interpret this as a causal impact of private provision.

As a result, our finding that short-term treatment impacts are higher in private courses is a descriptive statement, not a causal statement. In this section we extend beyond the pre-specified analysis to explore further what distinguishes these courses.

7.1 Comparing course and personal characteristics by provider type

We begin by comparing means of course, applicant, and labor market characteristics according to the whether or not the course is privately provided. Appendix Table 1 provides this comparison. Private courses are slightly longer than publicly provided courses, and are much more likely to be accounting courses and less likely to be craftsman, technical or service courses. They are less likely to be in Istanbul, and have somewhat less experienced teachers. Applicants are more educated for private courses than public courses, but are also slightly younger and have less previous work experience.

7.2 Propensity-score reweighting

To determine whether it is these observable differences between courses that are driving the greater effectiveness of vocational training in private courses we use the propensity-score reweighting approach of Hirano et al. (2003) and Nichols (2008). To do this, we estimate a propensity score for the likelihood of being a private course, restrict our sample to the common support of this propensity score, and then re-run equation (3) weighting public courses with weight $\frac{p(X)}{(1-p(X))}$, where $p(X)$ is the estimated propensity score. We use the characteristics compared in Appendix Table 1 to estimate this propensity score. This reweighting makes the private and public courses similar in terms of these characteristics.

As an alternative, we also estimate equation (3) controlling directly for the propensity score and its interaction with treatment status. The interaction of treatment with private provider then gives the differential treatment impact after controlling for any differential treatment impact associated with the propensity score.

Table 10 then compares the unweighted treatment interactions with private provider to the propensity-score weighted treatment interactions and to the estimates after controlling for the

propensity score and an interaction between treatment and propensity score. The point estimates are similar in magnitude across all three specifications and, generally retain their statistical significance. This suggests that the greater treatment impacts for private courses are not just due to observable differences in the characteristics of these courses and of the people applying for them. Instead, this evidence supports the theory that private providers, have better incentives to offer high quality courses more attuned to labor market needs. Nevertheless, as shown by the last column, this still does not appear to lead to lasting impacts on formal employment.

8. Cost-Benefit

The mean cost of a course is 1574 Turkish Lira (TL) per person, with privately provided courses having a mean cost of 1792 TL per person compared to 1455 TL per person for publicly provided courses. In addition to this, participants receive a stipend of 15 TL per day, which for an average course length of 57 days equates to a further 855 TL. The total cost to the government of providing a course thus averages 2429 TL per person (US$1619), and is 2692 TL (US$1795) per person for private courses.

Our LATE estimate of the overall gain in monthly income in Table 3 is 26 TL (although this is not statistically significant). If the course led to a sustained level increase in income of this amount, it would take 93 months for the gain in income to offset the cost of the course. However, our administrative data suggests any impacts on formal employment last at most 1.5 years, so that the costs exceed the benefits and the overall return to the course is negative.

Training was more effective for privately provided courses. The LATE estimate of the overall gain in monthly income for this group is a statistically significant 66 TL per month. It would take 41 months of these gains for the benefits to exceed the costs. However, since Table 9 shows that the formal employment impact of the course is no longer significant after 2.5 years, it also seems unlikely that the benefits to private courses exceed their costs.[19] As with all other vocational training evaluations in developing countries which we are aware of, we are unable to measure any general equilibrium spillover effects. Such effects have been detected by Crépon et al. (2013) for a job placement program for youth in France, who find some of the gains to treated

[19] These calculations are simplifications which ignore any opportunity cost of time or lost wages for participants during the time they take the course (since few were employed at baseline such costs are low); ignore any additional costs of distortions in the economy from raising the taxes to pay for these courses; and ignore any gains in utility associated with the non-wage benefits of being in more formal or higher quality jobs. These factors seem likely to be of second-order importance in our context.

youth come at the expense of the untreated. To the extent that trained individuals crowd out non-trained individuals from jobs, we are understating the cost per job created and overstating the returns. Turkey has approximately 2.5 million unemployed people. The 3,000 individuals trained in this experiment are thus a trivial fraction of all the unemployed, and even the 210,000 individuals trained in total in 2010 constitute less than 10 percent of the unemployed. This makes it seem likely that any such spillovers or crowding out may be second-order effects in our context.

9. Conclusions

This paper provides the first randomized evaluation of vocational training programs offered to the general unemployed population in a developing country setting. A large sample, containing both youth and older unemployed, as well as courses offered by the private and the public sector, enables us to examine the effectiveness of such training for a wider range of demographic characteristics and course types than existing literature. Linking participants to social security data enables us to trace the long-term trajectory of impacts on formal employment as well as avoid the attrition and measurement concerns that face existing studies in developing countries.

Our results show that the average impact of these vocational training programs is a very modest positive overall impact on employment and on the quality of employment, with the measured impact of the courses much less than expected by either course participants or government labor ministry staff. Despite the overall impact being close to zero, we find stronger and statistically significant impacts of vocational training in courses offered by private providers. Being selected for one of these courses results in a 4 to 6 percentage point increase in employment rates, relative to an employment rate of 37 percent for the control group. These returns persist when we control for observable differences in the characteristics of the courses and of their participants, but do not appear to last when it comes to impacts on formal employment over a 2.5 to 3 year period. Taken together the results suggest that there is some potential for vocational training to improve the short-term employment prospects of the unemployed, but that this potential will be best realized when courses are offered by providers that have both the incentives and ability to respond to market demands. However, overall the results suggest that this large-scale vocational training program struggles to meet a cost-benefit test. Given the renewed emphasis of these types of programs for many governments around the

world, these results suggest policymakers should be cautious in expecting such programs to have large impacts.

Appendix A: Definitions of Key Variables

Employment outcomes

The following employment outcomes are defined based on data collected in the follow-up survey.

Working at all: An indicator variable which takes value one if the individual has worked for cash or in-kind income at all in the past four weeks.

Employed 20 hours +: An indicator variable which takes value one if the individual is currently working for 20 hours a week or more.

Weekly hours: Hours worked per week in the last month employed. This is coded at 0 for individuals currently not working, and top-coded at 100 hours per week (the 99th percentile of the baseline response) to reduce the influence of outliers.

Monthly income: Total monthly income from work in the last month. This is coded as 0 for individuals not working, and top-coded at the 99th percentile of the control group earnings distribution (2500 Turkish Lira) to reduce the influence of outliers.

Transformed monthly income: The inverse hyperbolic sine transformation of monthly income from work in the last month, $log\left(y + (y^2 + 1)^{1/2}\right)$. This is intended to be more robust to outliers than levels of income and is similar to the logarithm transformation, but is also defined when income is zero (Burbidge et al, 1988).

Occupational status: this is coded based on work occupation using the international measures of socioeconomic occupational status of Ganzeboom and Treiman (1996). This is a continuous measure ranging from 16 (e.g. domestic helpers) to 90 (e.g. judges), and is coded as zero for individuals not working.

Formal work: this is an indicator variable coded as one if the individual is currently working in a job covered by social security.

Formal income: this is monthly income earned in jobs covered by social security.

Aggregate employment index: A standardized index obtained as the average of each of the above variables, after each has been standardized by subtracting the mean and dividing by its standard

deviation. This measure is set to missing for individuals who are missing data for the working at all variable, and otherwise is the average over all employment variables with non-missing data.

Ever formally employed between course end and January 2012: this is an indicator variable coded as one if the individual is reported in the social security administrative data as being in formal employment at any time between the end of their course date and January 2012, the midpoint of the follow-up survey.

Formally employed in August 2013: An indicator variable which takes the value one if social security administrative data reveal the individual to be formally employed in the month of August 2013.

Formal income earned in August 2013: Income for the month of August 2013, as reported in the social security administrative data.

Ever formally employed between course end and November 2013: : this is an indicator variable coded as one if the individual is reported in the social security administrative data as being in formal employment at any time between the end of their course date and November 2013.

Ever left a formal job between course end and November 2013: an indicator variable which is coded as one if the individual is reported in the social security administrative data as having left their first formal job worked in during this period.

Well-being measures

Expected probability of working in two years: The expected chance of having a job in two years time, coded as missing if an answer outside of the 0 to 100 range is given.

Mental health index MHI-5: this is a five item index of Veit and Ware (1983) which has a maximum score of 25 and minimum score of 5. Higher scores are desirable in that they indicate the experience of psychological well-being and the absence of psychological distress. Individuals are asked how often in the past four weeks they have done each of the following, each answered on a 5 point scale, where 1 denotes none of the time and 5 all of the time, and the MHI-5 is the sum of these responses:

- Been a nervous person (reverse-coded)
- Felt so down in the dumps that nothing could cheer them up (reverse-coded)
- Felt calm and peaceful
- Felt downhearted and blue (reverse-coded)
- Been a happy person

Current subjective well-being: individuals are asked which step on a 10-step Cantril ladder, where on the first step stand the poorest people, and on the tenth step standard the richest, they think their household stands today.

Subjective well-being in five years: Which step on the Cantril ladder they think their household will be on in 5 years time.

Total household income in past year: Income from all sources, top coded at the 99th percentile of the control group distribution(74,000 Turkish Lira)

Transformed household income: The inverse hyperbolic sine of total household income.

Durable asset index: the first principal component of 15 indicators of household durable asset and infrastructure ownership (own a gas or electric oven; own a microwave oven; own a dishwasher; own a DVD/VCD player; own a camera; have Digiturk/Satelite; own an air conditioner; own a CD player or iPod; own a telephone; own a computer; have an internet connection; own a private car; own a taxi, minibus or commercial vehicle; own a bicycle; have 4 or more rooms in their house).

Human capital measures

Is the main decision maker for where they work: based on a question which asks them who in the household makes decisions over whether they can work outside the home.

Raven test score: score out of 12 on a Raven Progressive Matrices test

Numerate: was able to answer four computational questions involving time, percentages, division and subtraction correctly.

Work centrality: Answer to the question "The most important thing that happens in life involves work", where 5 = strongly agree, and 1 = strongly disagree

Tenacity: this measures the extent to which individuals persist in difficult circumstances and is taken as the sum of responses on two questions taken from Baum and Locke (2004).

References

Abdulkadiroglu, Atila, Joshua Angrist, Susan Dynarski, Thomas Kane and Parag Pathak (2011) "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots." *Quarterly Journal of Economics* 126(2): 699-748.

Attanasio, Orazio, Adriana Kugler, Costas Meghir (2011) "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial", *American Economic Journal: Applied Economics* 3(3): 188-220.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman (2012) "Empowering Adolescent Girls: Evidence from a Randomized Control Trial in Uganda", Mimeo. LSE.

Baum, J. Robert and Edwin A. Locke (2004) "The Relationship of Entrepreneurial Traits, Skill, and Motivation to Subsequent Venture Growth", *Journal of Applied Psychology* 89(4): 587-98.

Benjamini, Yoav and Yosef Hochberg (1995) “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”, *Journal of the Royal Statistical Society Series B*, 57(1): 289-300.

Betcherman, Gordon, Karina Olivas, and Amit Dar (2004) “Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries.” World Bank Social Protection Discussion Paper Series 0402.

Blattman, Christopher, Nathan Fiala and Sebastian Martinez (2013) “Generating skilled self-employment in developing countries: Experimental evidence from Uganda””, *Quarterly Journal of Economics*, forthcoming.

Bruhn, Miriam and David McKenzie (2009), “In Pursuit of balance: Randomization in Practice in Development Field Experiments,” *American Economic Journal: Applied Economics* 1(4): 200-232.

Burbidge, John, Lonnie Magee and A. Leslie Robb (1988). "Alternative Transformations to Handle Extreme Values of the Dependent Variable" *Journal of the American Statistical Association* 83(401): 123-127

Card, David, Pablo Ibarraran, Ferdinando Regalia, David Rosas-Shady, and Yuri Soares (2011) “The Labor Market Impacts of Youth Training in the Dominican Republic”, *Journal of Labor Economics* 29(2): 267-300.

Casey, Katherine, Rachel Glennerster and Edward Miguel (2012) “Reshaping Institutions: Evidence on Aid Impacts using a Pre-analysis plan”, *Quarterly Journal of Economics* 127(4): 1755-1812.

Cho, Yoonyoung, Davie Kalomba, Mushfiq Mobarak and Victor Orozco (2013) “Gender Differences in the Effects of Vocational Training: Constraints on Women and Drop-Out Behavior”, *IZA Discussion Paper no. 7408.*

Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora (2013) “Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment”, *Quarterly Journal of Economics* 128(2): 531-80.

Delavande, Adeline, Xavier Gine and David McKenzie (2010) “Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence”, *Journal of Development Economics*, 94(2): 151-63

Fink, Günther, Margaret McConnell and Sebatian Vollmer (2012) “Testing for Heterogeneous Treatment Effects in Experimental Data: False Discovery Risks and Correction Procedures”, Mimeo. Harvard School of Public Health.

Ganzeboom, Harry and Donald Treiman (1996) “Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations”, *Social Science Research* 25(1): 201-39.

Groh, Matthew, Nandini Krishnan, David McKenzie and Tara Vishwanath (2012) “Soft skills or hard cash? The impact of training and wage subsidy programs on female youth employment in Jordan”, *World Bank Policy Research Working Paper* no. 6141.

Hanushek, Eric and Steven Ritkin (2006) “Teacher Quality”, pp. 1051-78 in E. Hanushek and F. Welch (eds.) *Handbook of the Economics of Education Volume 2*. Elsevier, Amsterdam.

Heckman, James J., Robert J. Lalonde and Jeffrey A. Smith (1999) “The Economics and Econometrics of Active Labor Market Programs”, *Handbook of Labor Economics Volume 3, Part A*, 1865-2097

Hirano, Keisuke, Guido Imbens and Geert Ridder (2003) “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score”, *Econometrica* 71(4): 1161-89.

International Labour Organization (ILO) (2012) *World of work report 2012: Better jobs for a better economy.* ILO, Geneva.

Jespersen, Svend, Jakob Munch and Lars Skipper (2008) “Costs and Benefits of Danish Active Labour Market Programmes”, *Labour Economics* 15(5): 859-84.

King, Kenneth and Robert Palmer (2010) *Planning for technical and vocational skills development*. UNESCO: International Institute for Educational Planning, Paris.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, ‘‘Experimental Analysis of Neighborhood Effects,’’ *Econometrica*, 75 (1): 83–119.

Kluve, Jochen (2010) “The effectiveness of European active labour market policy”, *Labour Economics* 17(6): 904-18.

Lechner, Michael and Conny Wunsch (2009) “Are Training Programs More Effective When Unemployment Is High?”, *Journal of Labor Economics* 27(4): 653-92.

Lee, David (2009) “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", *Review of Economic Studies* 76(3): 1071-1102.

Maitra, Pushkar and Subha Mani (2012) “Learning and Earning: Evidence from a Randomized Evaluation in India”, Mimeo. Department of Economics, Monash University.

McKenzie, David and David Robalino (2010) “Jobs and the Crisis: What has been done and where to go from here?”, *Viewpoint* note number 325, The World Bank Finance and Private Sector Development Vice Presidency.

Nichols, Austin (2008) “Erratum and discussion of propensity-score reweighting”, *Stata Journal* 8(4): 532-39.

Raaum, Oddbjørn, Hege Torp and Tao Zhang (2003) “Business cycles and the impact of labour market programmes”, Oslo University Department of Economics working paper 14/2002

Rosholm, Michael and Lars Skipper (2009) “Is Labour Market Training a Curse for the Unemployed? Evidence from a Social Experiment”, *Journal of Applied Econometrics*, 24 (2), 338-365

Schochet, Peter, John Burghardt and Sheena McConnell (2008) “Does Jobs Corp Work? Impact findings from the national Job Corps study”, *American Economic Review* 98(5): 1864-86.

UNESCO (2012) *Education for all Global Monitoring Report 2012: Youth and Skills – putting education to work.* UNESCO, Paris.

World Bank (2012a) *Skills, Not Just Diplomas: Managing Education for Results in Europe and Central Asia*. World Bank, Washington D.C.World Bank (2012b) *World Development Report 2013: Jobs*. World Bank, Washington D.C.

Figure 1: Geographic Distribution of Provinces

Provinces selected for experiment are indicated by orange rectangles around their names.

Figure 2: Trajectory of Formal Employment Impact

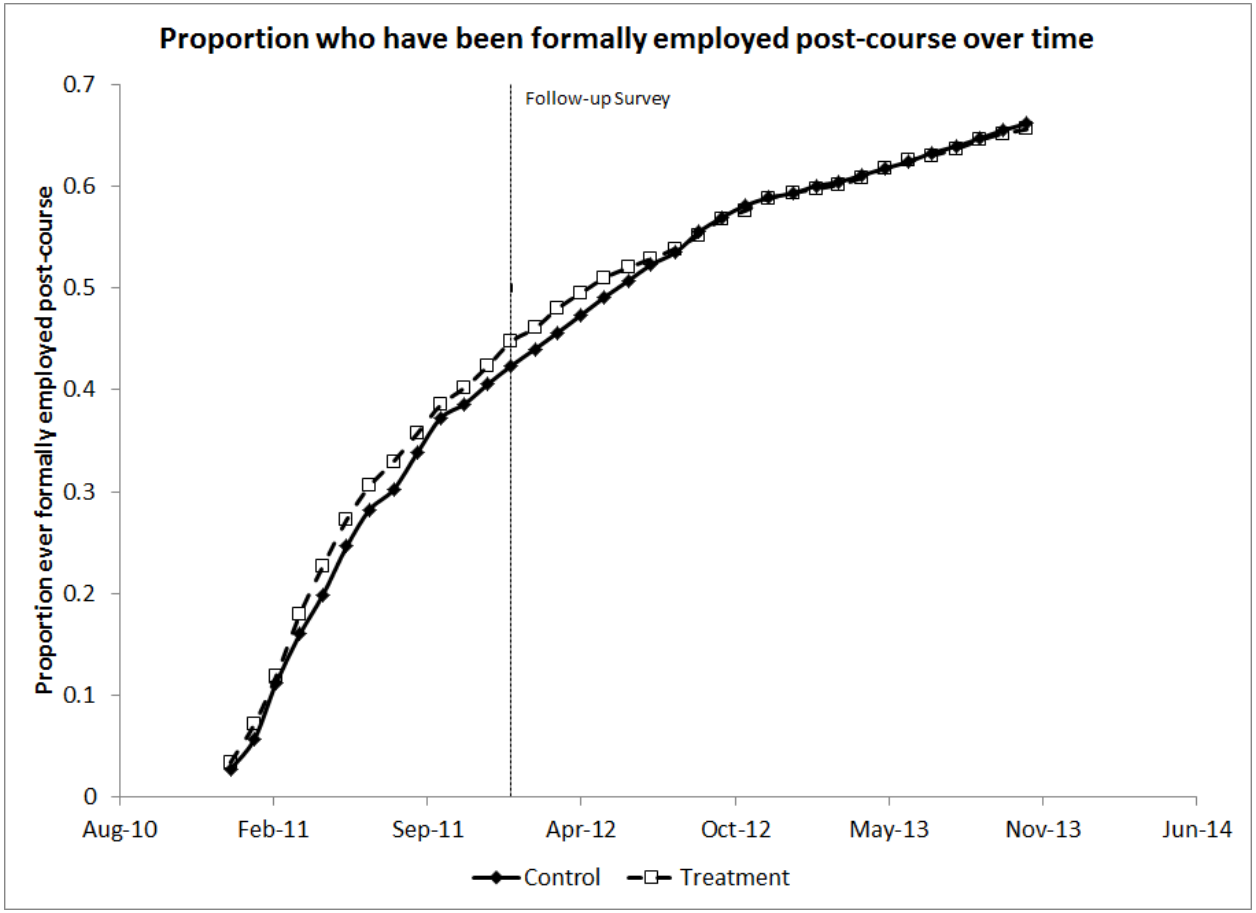

Table 1: Is attrition related to treatment status?

	Full Sample		Male Youth		Older Males		Female Youth		Older Females	
	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up	Baseline	Follow-up
Assigned to Treatment	-0.027***	-0.014**	-0.059***	-0.031**	-0.059***	-0.038**	-0.020	-0.001	0.001	-0.002
	(0.008)	(0.006)	(0.020)	(0.015)	(0.018)	(0.015)	(0.015)	(0.012)	(0.012)	(0.010)
Control group attrition rate	0.114	0.070	0.155	0.085	0.130	0.084	0.111	0.064	0.087	0.058
Sample size	5902	5902	1146	1146	1059	1059	1583	1583	2114	2114

Notes: Robust standard errors in parentheses. *, **, and *** indicate significance at the 10, 5 and 1 percent levels respectively.
All regressions include controls for course*gender*age group lotteries applied for.
Youth denotes individuals under 25, Older denotes individuals 25 and up.

Table 2: Balance and Summary Statistics

	Full Experimental Sample			Baseline Sample			Follow-up Sample		
	N	Control Mean (S.D.)	Treatment Difference (Std. Error)	N	Control Mean (S.D.)	Treatment Difference (Std. Error)	N	Control Mean (S.D.)	Treatment Difference (Std. Error)
Course Characteristics									
Course length (days)	5877	56.5	-0.005	5284	56.4	0.038	5507	56.4	-0.044
		(23.4)	(0.073)		(23.2)	(0.082)		(23.4)	(0.071)
Course length (hours)	5877	336	0.066	5284	334	0.246	5507	335	-0.241
		(151)	(0.453)		(150)	(0.506)		(151)	(0.439)
Private Provider	5899	0.349	0.000	5305	0.347	0.001	5526	0.351	0.000
		(0.477)	(0.002)		(0.476)	(0.002)		(0.477)	(0.002)
Accounting course	5877	0.242	0.000	5284	0.240	-0.000	5507	0.244	0.000
		(0.428)	(0.002)		(0.427)	(0.002)		(0.429)	(0.002)
Professional course	5877	0.122	-0.001	5284	0.116	-0.000	5507	0.120	-0.001
		(0.328)	(0.001)		(0.320)	(0.001)		(0.325)	(0.001)
Craftsman course	5877	0.156	-0.000	5284	0.155	-0.000	5507	0.158	-0.000
		(0.363)	(0.001)		(0.362)	(0.001)		(0.364)	(0.001)
Technical course	5877	0.147	0.000	5284	0.148	-0.000	5507	0.144	0.001
		(0.354)	(0.001)		(0.356)	(0.001)		(0.351)	(0.001)
Service course	5877	0.211	0.001	5284	0.217	0.000	5507	0.211	0.000
		(0.408)	(0.001)		(0.413)	(0.001)		(0.408)	(0.001)
Course in Istanbul	5902	0.220	-0.000	5308	0.217	-0.001	5529	0.211	0.000
		(0.414)	(0.001)		(0.412)	(0.001)		(0.408)	(0.001)
Individual Characteristics (administrative data)									
Female	5902	0.617	-0.000	5308	0.629	-0.002	5529	0.623	-0.001
		(0.486)	(0.002)		(0.483)	(0.002)		(0.485)	(0.002)
Age	5902	27.0	-0.041	5308	27.1	-0.089	5529	27.0	-0.058
		(7.2)	(0.109)		(7.2)	(0.116)		(7.2)	(0.113)
At least high school	5902	0.725	-0.004	5308	0.724	-0.005	5529	0.724	-0.004
		(0.447)	(0.009)		(0.447)	(0.009)		(0.447)	(0.009)
Individual characteristics (baseline data)									
Years of education				5255	11.3	-0.005	5008	11.3	0.014
					(3.3)	(0.069)		(3.3)	(0.071)
Has done previous training course				5308	0.264	-0.007	5057	0.265	-0.008
					(0.441)	(0.012)		(0.441)	(0.012)
Household head				5276	0.134	-0.000	5027	0.133	-0.003
					(0.340)	(0.008)		(0.340)	(0.008)
Household size				5308	4.09	0.024	5057	4.10	0.011
					(1.57)	(0.040)		(1.57)	(0.040)
Married				5033	0.346	0.004	4797	0.351	0.002
					(0.476)	(0.011)		(0.478)	(0.012)
Ever employed				5308	0.631	-0.021*	5057	0.626	-0.020
					(0.483)	(0.012)		(0.484)	(0.013)
Total years working for pay				5277	3.38	0.006	5027	3.33	0.006
					(4.91)	(0.111)		(4.88)	(0.113)

Note: Standard errors are robust standard errors. *, **, and *** indicate significance at the 10, 5 and 1 percent levels respectively. Standard errors based on regression with controls for course*gender*age group lotteries applied for. Sample sizes for the follow-up data for the individual characteristics measured at baseline are for the sample interviewed at both baseline and follow-up - there are also individuals interviewed at follow-up but not at baseline.

Table 3: Impact of Training on Employment Outcomes (Survey Data)

	Working at all	Employed 20 hours+	Weekly Hours	Monthly Income	Transformed Monthly Income	Occupational Status	Formal Work	Formal Income	Aggregate Employment Index
ITT Estimate	0.020	0.012	0.860	17.316	0.121	0.962*	0.020*	22.166*	0.039*
	(0.013)	(0.013)	(0.680)	(12.271)	(0.093)	(0.573)	(0.012)	(11.965)	(0.023)
Lee lower bound ITT	0.011	0.003	-0.103	-4.959	0.050	0.922	0.011	-0.231	0.016
	(0.013)	(0.013)	(0.668)	(11.720)	(0.093)	(0.573)	(0.012)	(11.378)	(0.023)
Lee upper bound ITT	0.026**	0.018	1.145*	22.592*	0.165*	1.242**	0.025**	26.460**	0.050**
	(0.013)	(0.013)	(0.684)	(12.331)	(0.094)	(0.576)	(0.012)	(12.034)	(0.023)
LATE Estimate	0.029	0.019	1.294	25.989	0.182	1.452*	0.030*	33.243*	0.059*
	(0.018)	(0.018)	(0.980)	(17.624)	(0.134)	(0.828)	(0.017)	(17.184)	(0.033)
Control Mean	0.420	0.361	17.922	299.109	2.541	17.128	0.293	257.887	0.001
Control Standard Deviation	0.494	0.480	25.545	464.600	3.516	21.763	0.455	448.160	0.871
Sample Size	5497	5529	5439	5396	5396	5418	5508	5464	5497

Notes: Robust standard errors in parentheses. *, **, and *** indicate significance at the 10, 5 and 1 percent levels respectively.
All regressions include controls for course*gender*age group lotteries applied for.
Appendix A defines the outcome variables.

Table 4: Impact of Training on Employment as Measured in Social Security Data

	Ever formally employed between course end and Jan 2012 (time of follow-up survey)	Formally employed in August 2013	Formal Income earned in August 2013	Ever formally employed between course end and Nov 2013	Ever left a formal job between course end and Nov 2013
ITT Estimate	0.026**	-0.009	-4.427	-0.005	0.015
	(0.012)	(0.012)	(19.935)	(0.011)	(0.012)
Control Mean	0.424	0.427	553.820	0.662	0.506
Control Std. Dev.	0.494	0.495	796.571	0.473	0.500
Number of Observations	5896	5896	5896	5896	5896

Notes: Robust standard errors in parentheses. *, **, and *** indicate significance at the 10, 5 and 1 percent levels respectively.
All regressions include controls for course*gender*age group lotteries applied for.

Table 5: Heterogeneity in Employment Outcomes by Age*Gender Stratifying Variables

					Transformed			
	Working at all	Employed 20 hours+	Weekly Hours	Monthly Income	Monthly Income	Occupational Status	Formal Work	Formal Income
ITT for males under 25	-0.021	-0.034	-0.952	-12.112	-0.241	-0.495	0.003	20.734
	(0.030)	(0.030)	(1.626)	(31.384)	(0.223)	(1.299)	(0.028)	(30.160)
	[-0.04, -0.01]	[-0.06*, -0.02]	[-2.5, -0.3]	[-58**, 1]	[-0.40*, -0.15]	[-1.4, +0.1]	[-0.02, 0.02]	[-24.9, 33.0]
ITT for males over 25	0.069** (a)	0.047	2.917*	35.304	0.383	2.997**(a)	0.047	32.019
	(0.029)	(0.031)	(1.772)	(37.501)	(0.244)	(1.332)	(0.032)	(36.733)
	[0.06*, 0.10***]	[0.03, 0.07**]	[1.0, 4.2**]	[-17, 62*]	[0.23, 0.58**]	[1.4, 4.3***]	[0.03, 0.07**]	[-28, +55]
ITT for females under 25	0.011	0.016	0.952	14.938	0.140	0.567	0.017	15.571
	(0.025)	(0.025)	(1.355)	(20.999)	(0.181)	(1.161)	(0.024)	(20.519)
	[0.01, 0.01]	[0.02, 0.02]	[0.8, 1.0]	[12, 15]	[0.13, 0.14]	[0.5, 0.6]	[0.02, 0.02]	[13, 16]
ITT for females over 25	0.023	0.018	0.761	26.052	0.174	1.034	0.019	23.058
	(0.020)	(0.019)	(1.001)	(17.321)	(0.141)	(0.917)	(0.018)	(16.788)
	[0.02, 0.03]	[0.02, 0.02]	[0.7, 0.8]	[22, 27]	[0.16, 0.18]	[1.0, 1.1]	[0.02, 0.02]	[19, 24]
p-value for testing equality:	0.188	0.276	0.457	0.712	0.268	0.294	0.768	0.982
Control Mean: young men	0.489	0.407	20.1	347	2.90	19.2	0.307	275
Control Mean: older men	0.621	0.548	27.3	518	3.94	24.3	0.466	458
Control Mean: young women	0.403	0.356	18.1	269	2.50	17.3	0.295	238
Control Mean: older women	0.295	0.247	12.0	187	1.69	12.4	0.196	161
Sample size	5497	5529	5439	5396	5396	5418	5508	5464

Notes: Robust standard errors in parentheses. *, **, and *** indicate significance at the 10, 5 and 1 percent levels respective
(a) denotes significant after applying the Benjamini-Hochberg procedure to maintain the false discovery rate across the four
Square brackets show Lee upper and lower bounds, along with indicator of their significance.
All regressions include controls for course*gender*age group lotteries applied for.
Test of equality tests the null of equality of treatment impact across the four age*gender strata.
Appendix A defines the outcome variables.

Table 6: Heterogeneity with respect to pre-specified individual characteristics

		Differential Impact on:			
Treatment Interaction with:	Control Mean (Std. Dev.) of Interacting Variable	Employed 20+ hours	Aggregate Employment Index	Ever formal by Jan 12	Currently formal Aug 13
Expected benefit from ISKUR training	31.9	0.000	0.001	-0.000	0.000
	(27.3)	(0.001)	(0.001)	(0.000)	(0.000)
Post-high school education	0.434	0.014	0.018	0.009	-0.017
	(0.496)	(0.027)	(0.050)	(0.027)	(0.027)
Previously taken part in a training course	0.264	0.062**	0.106*	0.030	0.033
	(0.441)	(0.031)	(0.057)	(0.031)	(0.031)
Has a child aged 6 or under	0.117	-0.047	-0.055	0.070*	-0.029
	(0.322)	(0.040)	(0.072)	(0.038)	(0.039)
Is the main decision-maker for whether they work	0.668	0.050*	0.088*	0.030	-0.041
	(0.471)	(0.028)	(0.050)	(0.028)	(0.028)
Raven test score	5.94	0.003	0.006	0.006	-0.004
	(3.10)	(0.004)	(0.008)	(0.004)	(0.004)
Numerate	0.564	0.003	0.027	0.013	-0.020
	(0.496)	(0.027)	(0.049)	(0.027)	(0.026)
Work centrality	4.08	0.023	0.045*	0.020	0.000
	(0.90)	(0.015)	(0.028)	(0.014)	(0.015)
Tenacity	8.37	0.010	0.018	0.010	-0.026***
	(1.30)	(0.011)	(0.019)	(0.010)	(0.010)
Unemployed for above the median duration	0.467	0.032	0.040	-0.004	0.015
	(0.499)	(0.027)	(0.049)	(0.027)	(0.027)

Notes: robust standard errors in parentheses, *, **, and *** indicate significance at the 10, 5 and 1 percent levels
Each row shows the treatment interaction from a regression which includes controls for course*age*gender lotteries, and for the level of the variable being interacted.

Table 7: Impact on Individual Well-Being and Household Well-Being

	Individual outcomes				**Household Outcomes**		
	Expected Prob. Of Working in 2 years	MHI-5 mental health (higher better)	Current Subjective Well Being	Subjective Well Being in 5 years	Household Income in last year	Transformed Household Income in last year	Durable Asset Index
ITT Estimate	0.923	-0.060	0.066*	0.061	485.524	0.057	0.106**
	(0.915)	(0.092)	(0.040)	(0.053)	(392.553)	(0.053)	(0.044)
Control Group Mean	54.1	18.508	4.436	5.822	21711	10.168	-0.071
Control Group Std. Dev.	33.4	3.410	1.514	1.980	14674	1.966	1.734
Sample Size	4878	5437	5508	5289	5396	5396	5495

Notes: Robust standard errors in parentheses. *, **, and *** indicate significance at the 10, 5 and 1 percent levels.
All regressions include controls for course*gender*age group lotteries applied for.
Appendix A defines the outcome variables.

Table 8: Comparison of Actual and Expected Treatment Impacts

	Employment Levels(%)		Treatment Impact (Proportion)		
	Expected percent chance of being employed if not trained (Std. dev)	Actual Control Employment Levels	ISKUR staff Expected Impact Mean (Std. dev)	Individual Expected Impact Mean (Std. dev)	Actual Impact LATE Estimate (std. error)
Full Sample	31.3	36.1	0.243	0.324	0.019
	(24.1)		(0.239)	(0.275)	(0.018)
Males under 25	35.4	40.7		0.297	-0.044
	(24.6)			(0.278)	(0.047)
Males 25 and over	33.5	54.8		0.302	0.081
	(26.1)			(0.280)	(0.054)
Females under 25	31.5	35.6		0.329	0.021
	(22.5)			(0.266)	(0.033)
Females 25 and over	27.7	24.7		0.346	0.022
	(23.5)			(0.277)	(0.027)

Notes: individual expectations are those of the treatment group at baseline.
Actual control employment level and LATE estimate for employment impact are for the outcome of working 20 hours or more per week.

Table 9: Which course characteristics are associated with better impacts?

Treatment Interaction with:	Sample Size	Control Mean (Std. Dev.) of Interacting Variable	Differential Impact on: Employed 20+ hours	Aggregate Employment Index	Ever formal by Jan 12	Currently formal Aug 13
Measures of course attendance rates						
Proportion assigned to course who attended it	5497	0.765	0.114	0.173	0.014	0.144
		(0.140)	(0.092)	(0.172)	(0.117)	(0.087)
Proportion assigned to course who completed it	5497	0.723	0.055	0.118	0.018	0.148*
		(0.164)	(0.083)	(0.155)	(0.095)	(0.081)
Proxies for course quality						
Course length above 320 hours	5494	0.423	-0.042*	-0.082*	-0.018	-0.024
		(0.494)	(0.024)	(0.044)	(0.027)	(0.023)
Average teacher experience greater than 12 months	4833	0.418	0.006	0.017	-0.017	0.030
		(0.493)	(0.026)	(0.048)	(0.029)	(0.025)
Percent of course teachers with tertiary education	4833	65.0	-0.000	-0.000	-0.000	-0.000
		(43.2)	(0.000)	(0.001)	(0.000)	(0.000)
Course has two or more competitors	4833	0.674	0.039	0.085*	0.042	0.004
		(0.468)	(0.026)	(0.049)	(0.030)	(0.028)
Course offered by private provider	5494	0.348	0.044*	0.117*** (a)	0.062**	0.025
		(0.477)	(0.023)	(0.043)	(0.028)	(0.025)
Measures of what trainees thought course did						
Proportion who thought course taught new technical skills	5497	0.796	-0.087	-0.209	-0.134	-0.049
		(0.127)	(0.093)	(0.185)	(0.100)	(0.108)
Proportion who thought course certified skills they had already	5497	0.842	-0.012	-0.096	-0.062	0.093
		(0.103)	(0.108)	(0.213)	(0.104)	(0.107)
Proportion who thought course taught new job finding strategies	5497	0.604	-0.025	-0.067	0.000	0.073
		(0.147)	(0.090)	(0.166)	(0.108)	(0.084)
Proportion who thought course made them aware of new jobs	5497	0.449	-0.052	-0.153	-0.154**	0.007
		(0.183)	(0.072)	(0.140)	(0.068)	(0.060)
Measure of labor market demand						
Provincial unemployment rate is above median	5497	0.463	0.013	0.033	-0.001	0.049**
		(0.499)	(0.023)	(0.044)	(0.026)	(0.023)
Type of course						
Accounting course	5497	0.242	0.059**	0.134**	0.046	0.034
		(0.428)	(0.027)	(0.053)	(0.031)	(0.028)
Computer course	5497	0.153	0.002	0.018	0.080**	-0.039
		(0.360)	(0.033)	(0.063)	(0.038)	(0.035)

Notes: robust standard errors in parentheses, clustered at the course level. *, **, and *** indicate significance at the 10, 5 and 1 percent levels respectively. (a) indicates significance after applying the Benjamini-Hochberg procedure to maintain the false discovery rate over the 12 different interactions at 10 percent.

Each row shows the treatment interaction from a regression which includes controls for course*age*gender lotteries, and for the level of the variable being interacted.

Table 10: Propensity-Score Reweighted Interactions with Private Courses

	Differential Impact on:			
Treatment Interaction with:	Employed 20+ hours	Aggregate Employment Index	Ever formal by Jan 12	Currently formal Aug 13
Course offered by private provider				
Unweighted (as reported in Table 8)	0.044*	0.117***	0.062**	0.025
	(0.023)	(0.043)	(0.028)	(0.025)
Propensity-score Reweighted	0.075**	0.171**	0.070*	0.059
	(0.035)	(0.066)	(0.036)	(0.042)
Controlling for interaction of treatment with propensity score	0.069**	0.150***	0.069**	0.034
	(0.031)	(0.057)	(0.033)	(0.035)

Notes: robust standard errors in parentheses, clustered at the course level.
*, **, and *** indicate significance at the 10, 5 and 1 percent levels respectively.
Propensity score calculated as a function of personal, course and labor market characteristics reported in Appendix Table 1
Each row shows the treatment interaction from a regression which includes controls for course*age*gender lotteries, and for the level effect of the course being offered by a private provider

Appendix 1: What is different about private courses?

	Public Provider	Private Provider
Course Characteristics		
Course length (days)	55	60***
Course length (hours)	328	353***
Accounting course	0.13	0.45***
Professional course	0.11	0.14***
Craftsman course	0.23	0.02***
Technical course	0.19	0.07***
Service course	0.24	0.14***
Course in Istanbul	0.27	0.12***
Average teacher experience greater than 12 months	0.45	0.34***
Percent of course teachers with tertiary education	71.9	51.2***
Individual Characteristics (administrative data)		
Female	0.61	0.66***
Age	27.0	26.5***
At least high school	0.66	0.85***
Individual characteristics (baseline data)		
Years of education	9.7	10.9***
Has done previous training course	0.25	0.20***
Household head	0.12	0.10**
Household size	3.86	3.64***
Married	0.30	0.27*
Ever employed	0.57	0.55
Total years working for pay	3.17	2.64***
Raven test score (out of 12)	6.10	5.93*
Numeracy score (out of 4)	3.31	3.51***
Tenacity	8.36	8.35
Labor market characteristics (Turkstat)		
Non-agricultural unemployment rate	12.5	12.6
Agricultural sector employment share	22.9	21.6***
Industrial sector employment share	29.0	27.5***
Service sector employment share	48.1	50.9***
Female labor force participation rate	29.5	28.4***
Male labor force participation rate	73.0	71.6***

Note: Sample means shown. Turkstat data are 2011 Labor Force Statistics at the NUTS2 level, www.turkstat.gov.tr [accessed August 2012]

*, **, and *** indicate statistically different from public providers at the 10, 5, and 1 percent levels respectively.