

# Short- and Longer-Run Impacts of Management Training: An Experiment in Tanzania\*

Yuki Higuchi

*(Nagoya City University, Nagoya, Japan)*

Edwin Paul Mhede

*(Ministry of Industry, Trade, and Investment, Dar es Salaam, Tanzania  
and National Graduate Institute for Policy Studies, Tokyo, Japan)*

Tetsushi Sonobe

*(National Graduate Institute for Policy Studies, Tokyo, Japan)*

May 2016

## Abstract

This paper reports the results of a randomized controlled trial of short-term training programs featuring *Kaizen* management for small garment manufacturers in Dar es Salaam, Tanzania. We collected baseline and follow-up data of the firms in treatment and control groups four times in the span of four years with negligible incidence of attrition. The data reveal that the training intervention improved significantly the management practices of the treated firms, and that the impact remained significant even three years after the intervention. Moreover, the initially insignificant impacts of the training on value added became significant three years later for a sub-group of the treated firms that received both classroom and on-site training components. The data and interview materials indicate that the training effect on business performance was reinforced by the treated firms' adaptive efforts to select useful management practices and to modify them to fit their business operation.

Keywords: Africa, Randomized controlled trial, Management training, Small and medium firms, Industrial cluster

JEL classification: L2, M1, O1

---

\* We would like to thank Keijiro Otsuka, Yasuyuki Todo, Takashi Kurosaki and Shinsuke Uchida for their helpful comments on the paper's earlier versions. This research was supported by the World Bank Japan PHRD Trust Fund No. TF096317, JSPS KAKENHI Grant Numbers 25101002 and 15H06540, and Grant-in-Aid for Research in Nagoya City University.

## Highlights

- We conducted a randomized trial of business training for small firms in Tanzania.
- Significant training impacts on business performance were found three years later.
- Treated firms adopted a number of management practices soon after the training.
- They stopped using some practices later and instead assimilated other practices.

## 1. Introduction

While it is entrepreneurs who play the central role in industrial development and job creation, their managerial capacity is severely limited in the developing world as Bruhn *et al.*, (2010), Bloom, Genakos, *et al.* (2012), and McKenzie and Woodruff (2015) among others attest. This is why there has been renewed interest in managerial capacity building through business training in recent years. An emerging body of literature finds, on the basis of randomized controlled trials, that training can improve management practices (e.g., Berge, Bjorvatn, Juniwaty *et al.*, 2012; Berge, Bjorvatn and Tungodden, 2014; Bruhn and Zia, 2013; De Mel *et al.*, 2014; Drexler *et al.*, 2014; Field *et al.*, 2010; Karlan *et al.*, 2012; Karlan and Valdivia, 2011; Mano *et al.*, 2012).

In their survey of this emerging literature, however, McKenzie and Woodruff (2014) point out that a further compilation of studies is needed to provide useful information for policy makers. Although the literature finds that training intervention induces treated firms to adopt good management practices, it has not presented clear evidence that training improves business performance in terms of profits, value added, or sales revenues. Such weak results could be attributed to noisy data, small sample sizes, or inadequately designed training programs, but it is also possible that the impacts were assessed too early. The existing studies tend to use data collected just a year after intervention, even though some business consultants suggest that it can take firms much longer time to see impacts on financial indicators than impacts on attitudes and practices (e.g., Imai, 2012; Morgan and Liker, 2006). It seems that only Bloom, Eifert *et al.* (2013) have explored this possibility so far.

The present paper asks what about if the period of follow-up observation is extended. Will we find a significant effect of training on business performance or an

even weaker effect? We conducted a baseline survey of randomly selected garment manufacturers in Dar es Salaam, Tanzania in early 2010 and then assigned them randomly to treatment and control groups. Our training program had classroom and on-site components. Classroom training participants learned from trainers about good management practices for about 50 hours in total. On-site training participants received several visits of the trainers to their workshops and concrete advice as to how to improve efficiency and safety at work. Follow-up surveys were conducted in 2011, 2012, and early 2014 to collect data of management practices at the moment and the annual values of production and costs in the previous calendar years. Thus, we extended the period of follow-up observation of business performance to three years.

Consistent with the existing studies mentioned above, this paper finds that the three treatment groups (i.e. those invited to one component or both training components) applied a significantly greater number of good practices soon after the training. It is also found that the training impacts on practices remained significant three years later, which is similar to the finding obtained by Higuchi *et al.* (2015) from a similar experiment in Vietnam. This paper reports another two major findings. First, while the treatment groups adopted a number of good practices at first, they later stopped using some of those practices because they found them not very useful. Second, however, participation in both components of the training program was found to have increasingly strong impacts on annual value added and sales revenue over time. These impacts became statistically significant at the latest survey.

One may wonder how a training program can improve the business performance of participants when they cease to use some of the practices that it taught. A possible explanation was given by some training participants, who told us that while they put

into practice as many training lessons as possible at first, they gradually discarded what was not very useful and adjusted the remaining practices to their own situations. This explanation is more plausible than other possible explanations, as we will discuss below.

The trainers were experienced providers of business development service (BDS) and familiar with the Start/Improve Your Business (SIYB) training materials developed by the International Labour Organization and widely used in developing countries.<sup>1</sup> Our training program, like the ones used in the existing studies of business training experiments, covered such standard subjects as entrepreneurship, marketing, and record keeping. Although our experiment concerns small enterprises, our sample enterprises are somewhat larger in employment size than microenterprises and microcredit clients, which are typical beneficiaries of above type of business training. For this reason, our program covered the issue of coordination among workers, featuring basic *Kaizen*, a common-sense approach to production management and quality control.<sup>2</sup>

Presumably because the sample enterprises were not very small, sample attrition is not serious in this study. Because the sample enterprises were drawn from a single industry in the same city, their heterogeneity to be controlled is relatively small compared with the studies whose samples are drawn from diverse industries. In the present study, however, information spillovers are beyond control. The training participants were so happy with the training program that they talked about it to their friends including those in the control group. Data on spillovers are presented below but not used in training assessment because of endogeneity: those participants who

---

<sup>1</sup> The information on SIYB training is available at <http://www.ilo.org/empent/areas/start-and-improve-your-business/lang--en/index.htm>

<sup>2</sup> Training in production management and quality control is not new to this literature: Bloom, Eifert *et al.* (2013) included these subjects in the training program for medium-sized textile firms in India.

anticipated good impacts of the training were less able to keep training contents secret. Since our impact assessment assumes away spillovers, our assessment results are expected to understate training impacts.

The remainder of the paper is organized as follows. Section 2 describes the experimental design. Section 3 addresses the issues of balance, attrition, and externality. Section 4 presents the impact evaluation results. Section 5 summarizes the findings and discusses implications for future studies.

## **2. Experimental Design**

### **2.1. Study Site and Sample Enterprises**

Dar es Salaam is a cluster of about 250 small-scale firms and numerous self-employed garment producers. The majority of these small-scale firms are owned and operated by female entrepreneurs. Presumably this is because of the sewing business training program provided by the United Nations Industrial Development Organization (UNIDO) provided in the 1990s to nurture female entrepreneurs. Some participants' sewing business became successful, which encouraged many other females and a smaller number of males in the city to start their own businesses. In this way, the cluster was established.

Our sample enterprises were selected from the member lists of three associations of business owners: Tanzania Handicraft Association (TANCRAFT), Handproducts of Tanzania (HOT) and Artisan Development Association of Tanzania (ADAT).<sup>3</sup> About 250 members of these associations sewed dresses, shirts, jackets, canvas tote bags, and

---

<sup>3</sup> There was another association of garment producers, but we did not include its members in the sample because they sold their products together in their own marketplace and acted as a cooperative sharing profits.

textile homeware, such as cushion covers and kitchen mittens. Out of these members, we selected 114 sample enterprises randomly for our survey.

## 2.2. Experimental Intervention

Most of the existing studies of management training interventions highlight basic training in accounting, marketing, and business strategy (McKenzie and Woodruff, 2014). Such training is suitable to microenterprises with zero to a few employees. Since our sample enterprises were somewhat larger in employment size, we included production management and quality control, which are mostly about coordination of the division of labor among workers, as a major subject of our training program. Among different approaches to this subject, *Kaizen* is probably in the most widespread use. It is a Japan-pioneered approach, based on US-born industrial engineering and quality management, emphasizing productivity improvement by the collaborative and continuous effort of entrepreneurs and their workers (Imai, 2012). In the 1980s, the US automobile industry adapted *Kaizen* to their circumstances and named their version *lean manufacturing*. Our training program featured the basic part of *Kaizen*.

We hired a Tanzanian lead trainer, who was qualified as an SIYB master trainer, and her three co-trainers. Since they had extensive experience in business training for small business owners and talked to participants in the local language, there is no reason to suspect that the trainers and participants had difficulty in communication. The trainers, however, were not familiar with *Kaizen*. We made a contract with a business consulting firm in Japan to send an experienced expert to our training program in Dar es Salaam. He communicated with the local trainers in English to transfer the essential knowledge of *Kaizen* and a method of teaching it to business owners and workers.

Although he taught training participants directly as well, the local trainers handled most part of the training program.

The timeline of the experiment is shown in Figure 1. We conducted a baseline survey of enterprises in April 2010, provided the classroom training in May and June 2010, and conducted a quick interim survey, which was followed by the on-site training and three full-scale follow-up surveys. The baseline survey collected detailed and reliable data from 109 enterprises. The classroom training took 2.5 hours a day, five days a week over a three-week span to teach the basic knowledge of *Kaizen* and standard contents of the SYIB training in entrepreneurship, marketing and record keeping.<sup>4</sup> To the classroom training, 52 randomly selected enterprises were invited, and 47 of them attended the training at least 10 days out of 15 days (three weeks). We regard these 47 enterprises as participants in this training component.

Between the classroom and on-site training components, the *Kaizen* expert and the local trainers worked on building two model workshops. Improvement in efficiency and safety sometimes requires a drastic change in workshop layout. Model workshops were intended to be showcase of streamlined layout. Moreover, the local trainers needed to learn from the *Kaizen* expert how to help enterprises change layout appropriately while building model workshops under his guidance. From among the classroom training participants, two enterprises were selected to be model workshops because they were eager to learn directly from the expert and willing to receive visits from other training participants who wanted to see their layouts and *Kaizen* practices and because the expert found their workshops suited for the training of trainers. Since

---

<sup>4</sup> The classroom training was supplemented by a one-week optional course on color coordination and product design.



these two model workshops received special treatments, they are excluded from our sample in the analysis below.

From among the remaining 107 enterprises, 54 enterprises were randomly selected to be invited to the on-site training, which began with a half-day seminar. All the invited owner/operators (hereafter referred to as entrepreneurs) attended the seminar and were randomly assigned to trainers. Each trainer visited to the workshops in his or her charge to assess the situations and make a coaching plan and, if necessary, a layout change plan, for each workshop. Later the trainer visited each workshop at least twice more and in many cases several times more. The number of follow-up visits was determined by the local consultants, depending on the willingness and availability of each trainee.

### 2.3. Randomization and Take-up

We made the random assignment of sample enterprises to the classroom and on-site components separately. As a result, we had four groups of sample enterprises: those who were invited to both the classroom and on-site components (and labeled as Group TT); those invited to the classroom training only (Group TC); those invited to the on-site training only (Group CT); and those invited neither component (Group CC). Although the assignment is not perfectly random because of the two enterprises that served as model workshops, it was intended to compare the effectiveness of classroom training and on-site training and to know whether either one is enough or whether both are needed to improve business performance. Compos *et al.* (2014) suggests a similar way of impact evaluation.

As mentioned earlier, 47 entrepreneurs participated in the classroom training

while 52 were invited. Thus, the take-up rate for the classroom training was 90 percent. The take-up rate for the on-site training was 100 percent. A possible reason for the higher take-up rate of the on-site training is that participants did not have to go to a training venue. Another possible reason is that sample entrepreneurs were more willing to receive training after the classroom training established good reputation of our program, part of spillover effects which we will discuss shortly.

All participants in the classroom and on-site training components were those invited to the respective component. There were no defiers who were not invited but participated in either or both training components.

### **3. Internal Validity**

#### **3.1. Balance**

During the baseline and follow-up surveys, our enumerators visited sample enterprises and conducted personal interviews with their entrepreneurs to gather information on their background characteristics, management practices, and business performance. Columns 1 to 4 of Table 1 present data on the characteristics by treatment status (on the invitation, not participation, basis) and indicate that average entrepreneurs are in their mid-40s, and that female entrepreneurs are dominant. A relatively wealthy ethnic group, Chagga, accounts for a quarter of our sample. The educational attainment of our sample entrepreneurs is much higher than the average schooling attainment in Tanzania, which is 5.1 years for those aged 25 or above as of 2010 according to Barro and Lee (2012). If male and female entrepreneurs are compared, females tend to have higher education attainment. Those entrepreneurs who had business training experience in the past accounts for more than 60 percent of the whole sample and an even higher

percentage of the female subsample. Only a small number of entrepreneurs had prior work experience in sewn product manufacturing.

Columns 5 to 7 of the table report the  $p$ -values for the tests of equality in the mean between each treatment group and the control group. Since the small sample is subdivided into four groups, there are two marginally significant differences and a five percent significant difference between groups. Otherwise, randomization was successful as far as the observable characteristics are concerned.

The management practices of the sample enterprises were evaluated by our enumerators based on visual inspection or the way that the entrepreneurs responded to their questions or both. The enumerators used 27 diagnostic criteria, which is listed in Appendix Table A1. Evaluation results are expressed by a management score, which is the number of the diagnostic criteria that the enumerator finds satisfied. Hence, the score ranges from zero to 27. As indicators of business performance, we use annual values of value added and sales revenue in international dollar.<sup>5</sup> Appendix 1 gives details of our baseline and follow-up surveys, construction of data, and our enumerators. Appendix 2 asks the question of how the marketing score, business performance measures, and enterprise characteristics are related. Appendix Table A2 reports the results of the regressions summarizing the associations among them.

Table 2 presents the management score and the business performance measures. We have four data points for the management score as we conducted baseline and three follow-up surveys, and five data points for value added and sales revenue as we collected recall annual data on these indicators of business performance. Because of

---

<sup>5</sup> We also used real values obtained by applying the GDP deflator. The results are available upon request to the corresponding author, but they are almost the same as the results reported here.

the randomization, the four groups are expected to share similar baseline values of these variables. This is generally the case, as shown in columns 5 to 7, even though there is a marginally significant difference in the baseline management score between Groups TT and CC.

### 3.2. Attrition

Except for the two enterprises that volunteered as model workshops, there was no incidence of sample attrition in the first follow-up survey in April 2011. When the second follow-up survey was conducted in September 2012, three sample enterprises (one in Group TC and two in Group CT) were closed, and their data were not collected. The missing enterprise in Group TC and one of the two missing enterprises in Group CT, however, were in operation at the time of the third follow-up survey in January 2014. Instead, this survey could not collect data of six sample enterprises (one in Group TT, three in Group CT, and two in Group CC).

The attrition rate in this study is lower than the attrition rates reported by many other studies of management training experiments (e.g., McKenzie and Woodruff, 2014). Presumably this may be attributed to the relatively large firm size of our sample enterprises in view of the vast literature on firm survival (e.g., Evans 1987; Söderbom *et al.* 2006). Although not reported in a table in this paper, we ran regressions to find that attrition is not correlated with three dummy variables indicating treatment status.

### 3.3. Externality

All sample enterprises are located in Dar es Salaam. Because of their geographical proximity, knowledge could easily spill over from training participants to

non-participants.<sup>6</sup> During the third follow-up survey, we showed the list of the sample entrepreneurs, together with their enterprise names and addresses, to each sample entrepreneur and asked with whom in the list he or she has talked about training contents. Row (a) of Table 3 summarizes the answers to this question. An entrepreneur belonging to one of the three treatment groups has discussed training contents with about 10 sample entrepreneurs on average. Row (b) reports the average number of participants in either or both components that a sample entrepreneur talked with about training contents.

In columns 1 and 2, the gap between (a) and (b) is small, suggesting that classroom training participants talked about training contents mostly with other participants rather than entrepreneurs belonging to the control group. As column 4 in row (b) shows, however, an entrepreneur in the control group also talked about training contents with more than three treatment group entrepreneurs. Less direct evidence for knowledge spillovers is that the management score of Group CC increased from 9.7 in the baseline survey to 12.4 in the first follow-up survey and 17.0 in the second follow-up survey (see column 4 of Table 2).

Another piece of suggestive evidence for knowledge spillovers come from our data on willingness to pay. We asked each sample entrepreneur if he or she was willing to pay the local currency equivalent of 150 USD for the training program after explaining the purpose and features of the program.<sup>7</sup> Row (c) shows that the

---

<sup>6</sup> De Grip and Sauermann (2012) find that knowledge spilt over from trained workers to co-workers. Bloom, Eifert *et al.* (2013) find that management knowledge spilt over from a treated factory to the other factories owned by the same entrepreneur.

<sup>7</sup> The question on willingness to pay is prone to reporting bias because of its hypothetical nature. In order to reduce such bias, we followed the lead of Blumenschein *et al.* (2008). After outlining our training program, we asked them a hypothetical question, “Would you pay 400,000 Tanzanian shilling (= 150 USD) to participate in the training program?” which was followed, if the answer

proportion of the sample entrepreneurs whose answer was affirmative. This proportion increased to 100 percent after the training program in all groups including Group CC. The increased willingness to pay among the control group entrepreneurs indicates that knowledge spilt over from training participants to them.

While the data lend support to the hypothesis that knowledge spillovers existed, the data do not allow us to control for their effects. This is because spillovers are endogenous to training impacts. It is easy to imagine those participants who anticipated good effects of training participation were so happy and more likely to be unable to keep the lessons that they learned from the training secret from non-participants including the control group entrepreneurs. Thus, we estimate training impacts based on the assumption that the control group was not affected by the training, i.e., the Stable Unit Treatment Value Assumption (SUTVA).

If, however, one ignores knowledge spillovers and assesses training impacts under SUTVA when spillovers exist and are not neutralized by opposite effects, the resulting assessment understates the impact (e.g., Rubin, 1978). A possible opposite effect of management training is for the treatment group to steal business from the control group (e.g., Bloom, Schankerman *et al.* 2013). For two reasons, we expect that business stealing if any is weak in the industry under study. First, markets for sewn products in East Africa have been increasingly integrated. Indeed, 60 percent of the sample enterprises were exporting their products to the neighboring countries and beyond, but similar products were imported and sold in large quantities in the city. In this situation, it seems unlikely that their sales increases systematically reduce the

---

was affirmative, by an additional question, “How sure are you about the answer? Are you definitely sure or probably sure?” Based on these set of questions, we constructed a dummy variable that is one if the answer is “definitely sure of the willingness to pay” and zero otherwise.

control group's sales revenues.

Second, our training program featured the *Kaizen* approach to coordinating the division of labor and reducing wasteful use of intermediate inputs as well as time. Although marketing is covered by the classroom component of the program, the emphasis of the training was placed on waste reduction rather than sales promotion. As we will see below, the estimate of the training impact on value added is more significant than that on sales revenue, probably reflecting the training program's emphasis on reducing the wasteful use of materials. Thus, it is more likely than not that our estimates obtained below understate the true training impacts.

#### **4. Empirical Analyses**

##### 4.1. Descriptive Analysis

As shown in Table 2, every treatment group had a higher average management score after the training program than before, but Group CC had a smaller increase in the score than the three treatment groups. The increase in the average score of Groups CC suggests that this group learned from training participants about good practices. Although we cannot completely deny the possibility that Group CC pretended to have adopted good practices, this possibility is low because the newly adopted practices by Group CC include those visually checked by the enumerators.

As shown in columns 5 to 7, Groups TT and TC exceeded Group CC significantly soon after the program, and 1.5 years later, Group CT also exceeded Group CC significantly. Even three years later, the differences in the management score between the three treatment groups and the control group remained significant. Thus, the management training improves management practices almost immediately as many

studies have already found (McKenzie and Woodruff, 2014), and the favorable effect is long-lasting as a similar study in Vietnam has found (Higuchi *et al.* 2015).

A new finding as to the training effect on management practices is that every group had a lower management score three years in the third follow-up survey than in the second follow-up survey. A possible explanation for the general decrease in management scores is that the enterprises stopped using those practices which they had adopted and found not very useful. Indeed, some sample entrepreneurs explained to us that their scores decreased in this way.

Turning to business performance, we find that all groups had initial increases and subsequent decreases in value added and sales revenues, and that the fluctuations were large relative to the respective initial amounts. If we take a closer look, we recognize that the ratio of value added to sales revenue gradually declined from about 60 percent in the baseline to about 50 percent in 2013, which our respondents attribute to a general increase in material prices coupled with a flood of low-priced imports since 2011. For all groups, while sales revenues in 2013 were greater than those in the baseline, value added returned to the baseline levels in 2013 for Groups TC, CT, and CC. Unlike these groups, Group TT increased value added per enterprise by 30 percent during this period. As a result, the difference between Groups TT and CC in mean value added became significant at the 5 percent level in 2013. Note that Group TT also increased median value added by 20 percent, while the other groups could not increase it.

To compare the four groups in terms of the entire distribution of value added, Figure 2 plots the cumulative distribution function (CDF) of value added by group in the baseline (Panel A) and in 2013 (Panel B). It appears from Panel A that the baseline CDF for Group TT is located to the right of those for the other groups except for the



largest enterprises. Indeed Group TT had the largest median value added among the four groups in the baseline, according to Table 2. Still the Kolmogorov-Smirnov test does not reject the null hypothesis of equality in the baseline distributions between Groups TT and CC. From Panel B, we see Groups TC and CT sharing a similar CDF with Group CC except for the largest enterprises of Group CT, which increased value added substantially. By contrast, the CDF of Group TT is almost entirely located to the right of those of the other groups, implying a near first-order stochastic dominance. The Kolmogorov-Smirnov test rejects the equal distribution hypothesis at the 1 percent significance level for Groups TT and CC in 2013.

Table 2 reports that the difference between Groups TT and CC is significant at the 5 percent level in not just value added but also sales revenue. Another point to note is that Group TT was significantly more likely to undertake fixed capital investment than Group CC in the period July 2012 - December 2013 (see two lines at the bottom of the table). A marginally significant difference is also found between Groups TC and CC. Data on investments are based on retrospective information collected in the second and third follow-up surveys. Since the number of enterprises that made investment is not large and the amount of reported investment varies greatly, we use a dummy variable indicating whether or not investment was undertaken. The Group TT's higher propensity to invest may have something to do with its higher business performance.

#### 4.2. Econometric Specification

To evaluate the impacts of the training on the management score and business performance, we consider the following regression equation:

$$y_{it} = \alpha + \beta_{Bt}Z_{Bi} + \beta_{Et}Z_{Ei} + \gamma y_{i0} + \sum_{n=1}^{N-1} \delta_n m_{nit} + \eta_t + \varepsilon_{it}, \quad (1)$$

where  $y_{it}$  is the management score of enterprise  $i$  at the  $t$ -th round of follow-up survey ( $t = 1, 2, 3$ ) or the enterprise's value added (or sales revenue) in year  $t$  ( $t = 2010, 2011, 2012, 2013$ ),  $Z_{Bi}$  is a dummy variable indicating whether enterprise  $i$  was invited to both components of the training program (i.e., whether the enterprise belongs to Group TT) or not, and similarly  $Z_{Ei}$  is a dummy variable indicating whether the enterprise was invited to either one of the two components (i.e., whether it belongs to Group TC or CT) or not.<sup>8</sup> Since we expect training effects to change over time, the coefficients on these variables,  $\beta_{Bt}$  and  $\beta_{Et}$ , have subscript  $t$ . In other words, we estimate three coefficients on  $Z_{Bi}$  and three on  $Z_{Ei}$  if the dependent variable is the management score and four coefficients if the dependent variable is value added (or sales revenue). The training effects that these coefficients are intended to capture are the intention-to-treat (ITT) effects of the two components of the training program and those of either one component.

To estimate these coefficients, we employ ANCOVA estimator, which is more efficient than the fixed-effect model estimator according to McKenzie (2012) and subsequent studies. Specifically, the right-hand side of equation (1) includes the mean,  $y_{i0}$ , of baseline values (i.e., values in 2008 and 2009 in the cases in which  $y$  stands for value added and sales revenue) of the dependent variable.

Equation (1) also includes a set of dummy variables,  $m_{nit}$ , indicating which of the  $N$  enumerators was in charge of collecting data from enterprise  $i$  in survey round  $t$ , to

---

<sup>8</sup> Groups TC and CT are combined because they share similar cumulative density functions as shown in Figure 2. If  $Z_{Ei}$  is split into two dummy variables indicating whether the enterprises was invited to the classroom training or not and whether it was invited to the on-site training or not, respectively, we obtained qualitatively the same results, which are available upon request.

control for enumerator fixed effects,  $\delta_n$ , i.e., heterogeneity in the way they evaluate management practices and elicit accounting information. Since eight enumerators were engaged in the three follow-up surveys (i.e.,  $N = 8$ ), equation (1) has seven enumerator dummies. Time effects common to all enterprises,  $\eta_t$ , is captured by two time dummies for management score and three time dummies for value added (or sales revenue). Error term,  $\varepsilon_{it}$ , is clustered to control for autocorrelation within respective enterprises.<sup>9</sup>

In addition to the ITT effects, we estimate the treatment on the treated (TOT) effects on the compliers who participated in the training program entirely or partially depending the randomly assignment, following the lead of Imbens and Angrist (1994). For this purpose, we instrument a dummy variable that indicates whether enterprise  $i$  participated in both components of the training program with the corresponding random assignment variable, i.e.,  $Z_{Bi}$ . Similarly,  $Z_{Ei}$  is used as an instrumental variable for a dummy variable that indicates whether enterprise  $i$  participated in only one component. We estimate a total of six coefficients on these two participation variables. As mentioned earlier, the compliance rate for the training programs was high. Thus, the estimates of TOT effects would not differ much from the ITT estimates, but the former are expected to be somewhat greater than the latter in magnitude.

#### 4.3. Estimation Results

Tables 4 and 5 present the estimates of equation (1) for the management score and business performance, respectively. Table 6 presents estimated impacts on investment.

---

<sup>9</sup> Equation (1) could include a set of variables characterising entrepreneurs that appear in Table 1 and Appendix Table A2. We tried this specification and found the estimation results differ little from the results of estimation without using these variables as reported in Tables 4 to 6.

In Table 4, columns 1 and 2 report the ITT and TOT estimates, respectively, for a total score, while columns 3 to 7 reports TOT estimates for scores by subject. The large *F*-statistics, reported in columns 2 to 7, indicate that our use of random assignment as the instrument variable is valid. Although the estimated coefficients on the enumerator dummies are not reported in the table, the null hypothesis that their true values are all zero is rejected at the 1 percent significance level, indicating that the inclusion of these dummies reduces estimation bias.

In columns 1 and 2, ITT and TOT estimates are similar, but TOT is somewhat greater as expected. According to column 2, which is our preferred specification, participation in both training components increases the total management score by 5.90 points soon after the training program, by 2.08 points 1.5 years later, and by 3.71 points three years later, compared with the control group. Participation in only one component increases the score by 2.74 points soon after the training, by 1.81 and by 3.48 points three years later. These estimated impacts of training participation on the management score are highly significant, and the increases in the score are indeed large compared with the initial score as shown at the bottom of the table. Of the two time dummies, the first one has a highly significant coefficient but the second one does not, indicating that the average management score of Group CC went up significantly in the first 1.5 years after the training but then it went down.

In columns 3 to 6, we see similar patterns: the sales promotion, record keeping, and *Kaizen* scores increased as a result of training participation; their increases were large soon after the training and three years later; and they went up 1.5 years later, according to the coefficient on the first time dummy. The marketing score is somewhat exceptional. In column 5, fewer coefficients on the training dummies are

significant, and the coefficient on the second time dummies is greater than that on the first one and both are highly significant.

These results shown in Table 4 suggest that several training lessons were soon applied in practice by training participants and then spilt over to non-participants, and that spillovers or imitation reduced the measured impacts on the training participants 1.5 years later. By the time of the latest survey, however, non-participants lost interest in the new practices of sales promotion, record keeping, and *Kaizen*, more than the participants. A possible explanation is that the training participants, especially those in Group TT, had better ideas to apply training lessons in practice or were able to make more profits from them than non-participants. Similarly, the reason why training participation made little difference in the marketing score may be that non-participants could find new practices of marketing useful or profitable and continued to use them.

Table 4, together with Table 2, indicates that the training participants and even non-participants kept using various management practices, not just a few particular practices, taught in the training practice three years after the training. This finding points to the usefulness of including different subjects in training programs. Since our sample enterprises tended to employ five workers or so, we included *Kaizen* training in our program so as to improve coordination of the division of labor among workers. Its inclusion has proved rewarding as column 6 of Table 4 attests.

Table 5 presents the estimates of Equation (1) for value added and sales revenue. Here, the baseline value  $y_{i0}$  is the mean of value added (or sales revenue) in 2008 and 2009, and there are four periods (i.e.,  $t = 2010, 2011, 2012$  and  $2013$ ). Columns 1 to 3 report the impact of the training on value added, and columns 4 to 6 report the impacts on sales revenue. While the dependent variable is the international dollar amount in

columns 1, 2, 4, and 5, it is transformed by means of the method developed by Burbidge et al. (1988) in columns 3 and 6. This transformation allows better handling of extreme values like logarithms, but it can be applied to negative values unlike logarithm.<sup>10</sup> While columns 1 and 4 report ITT estimates, the other columns report TOT estimates, together with the first-stage  $F$ -statistics, which are so large that the random assignment variables are valid instrument variables.

In this table, no explanatory variables have significant coefficients other than the interaction term, Both training  $\times$  the 2013 year dummy, whose coefficient is significant at the five percent level in columns (1) and (2) and at the ten percent level in columns (3) to (5). Thus, participation in both classroom and on-site training components has finally proved to have impacts on the levels of value added and sales revenue three years after the training. The estimated 7,512 dollar increase in value added due to training participation (see column 2) amounts to be more than half of the corresponding baseline level. As shown in columns 3 and 6, the application of the logarithm-like transformation makes the estimated training impacts on value added and sales revenue appear less significant. Presumably, this is because the transformation mutes the influence of the largest enterprises in Group TT, which tended to feel the strongest impacts as shown by the CDF's in Figure 2. The estimated training impacts on value added are more significant than those on sales revenue. Presumably, this is because the *Kaizen* training emphasized the importance of reducing material wastes so that incomes for entrepreneurs and workers would increase given production and sales revenue.

---

<sup>10</sup> In some cases, value added is negative largely because of wasted materials. The transformation is given by  $\log\{y + (y^2 + 1)^{0.5}\}$ .

Table 6 the estimated impacts of the training on investment in equipment. Our second and third follow-up surveys collected recall data on investment for two overlapping periods with different intervals, as shown in the table. The first two columns concern the earlier period, and the last two columns concern the later period. Columns 1 and 3 show the marginal-effect estimates of a probit model, to answer the question of whether the treated enterprises were more likely to undertake investment than the untreated enterprises, regardless of the amount of investment. Columns 2 and 4 report the OLS estimates, to answer the question of how much more the treated enterprises invested than the untreated ones. We see significant estimates only in column 3, which suggests that the training program increased investment in small equipment.

The results shown in Tables 4 to 6 are inconsistent with the hypothesis that the training effect on business performance inevitably diminishes over time. On the contrary, they are consistent with the hypothesis that it takes longer time to detect the effect of management training on business performance than that the effect on management practices. The delay may be caused by machine breaks, workers' injuries, transaction partners' breach of contract, and other idiosyncratic events that, together with management practices, affect business performance. Another possible explanation for the delay may be that it takes time before an enterprise assimilates good practices that it has adopted.

As Tables 2 and 3 suggest, it is likely that there were considerable spillovers of knowledge from the treatment group to the control group, which would attenuate the estimated impacts of the training on practices and performance considerably. Appendix 3 presents suggestive evidence for spillovers, reinforcing this view. Still,

three years after the training program, as Tables 4 and 5 indicate, the control group stopped using more practices than the treatment group, and the estimated impact of the training on value added became significant. A possible interpretation is that the treatment group had better assimilation of good practices than the control group.

## **5. Conclusions**

This paper has reported our follow-up observation for three years after a randomized controlled trial of management training for small enterprises. In line with the stylized finding of the existing studies, we have found that a short-term management training program improves management practices almost immediately. While training impact on business performance is found weak or insignificant in many of the existing studies, this paper has found that it became statistically significant three years after the intervention. Although our follow-up observation is still in progress, our tentative conclusion is that those existing studies evaluated management training impacts too early to detect a favorable impact on business performance.

There seem to be multiple reasons why a training impact on performance is weak or insignificant soon after management training and why it can become significant later. This paper, however, has not quantified the relative importance of those reasons. Nor has it offered firm evidence that knowledge spillovers attenuate the estimated impacts of training on practices and performance. It has provided evidence that many participants talked to non-participants about training contents, but we do not know whether their conversations really transferred useful knowledge to non-participants. These issues are left to future research. In addition, although our training program included production management and quality control as major training subjects, our experiment



design does not allow us to assess the impact of the inclusion of these subjects. Our findings merely suggest that different participants benefited from different training subjects. These knowledge gaps warrant considerable further compilation of experimental studies of management training with long periods of follow-up observation.

## **Appendix 1: Surveys**

Our enumerators conducted personal interviews with entrepreneurs while filling questionnaires. They kept using basically the same questionnaires. We deleted some questions that turned out to be meaningless or uninteresting. Among such questions, the most important one is about the willingness to pay for training participation (see Table 3). At the interim survey in 2010, almost all sample entrepreneurs, regardless of treatment status, answered affirmatively. At the first follow-up survey, all of them answered affirmatively. Thus we decided not to ask the questions about this issue subsequently. This was only one major addition to the questionnaire. At the interim survey, we added a new section on acquaintances and conversations about training contents (see Appendix 3 below), which were not included in the questionnaire used in the baseline survey, and we kept expanding this section by adding questions about visiting other enterprises and copying practices or layouts of other enterprises.

Our questions about management practices are about the 27 diagnostic criteria as listed in Appendix Table A1. For many criteria, visual inspection was used to judge whether criteria were satisfied or not. For criteria like “The entrepreneur has a clear sales or profit target in this year,” visual inspection is impossible, and direct questions like “do you have a clear sales or profit target?” may lead to false answers. The questionnaire instead asks the entrepreneur to talk about their sales or profit target, and it asks the enumerator to judge whether the entrepreneur’s target is clear. In this way, we tried to construct the management score.

We measure business performance by value added and sales revenue. The questionnaire does not directly ask value added, even though de Mel *et al.* (2009) recommend asking entrepreneurs directly how much profits they earn, rather than

calculating profits. Our enumerators asked the entrepreneur the quantity and price of each of their products, those of each material, subcontracting costs, energy costs, and transportation and communication costs. Written records were also used whenever available. The enumerators then estimated value added carefully and showed the estimate to the entrepreneur. If the estimate did not make sense to the entrepreneur, the enumerator would elicit further information to revise the estimate of value added until the amount made sense to him or her. As enumerators require capacity for such data collection procedure, all enumerators held a bachelor's degree and received survey training from us before each of the baseline and follow-up surveys.

## **Appendix 2: Baseline Correlations**

The baseline management score is closely associated with the number of years of schooling and the number of years of business operation and weakly associated with past business training experience. See Appendix Table A2. These results are consistent with the findings of Bloom and van Reenen (2010) that the management practice score is closely associated with the human capital of managers.

Baseline value added and sales revenue are closely correlated with the baseline management score as well as with schooling, past business training experience, and business operation experience. But these business performance variables and business operation experience remain closely associated even when the baseline management score is controlled for.

## **Appendix 3: Knowledge Spillovers**

Appendix Table A3 reports the results of estimation that incorporates the number of

contacts on the right-hand side of equation (1). Our data set has two contact variables: one is the number of the treated entrepreneurs whom the respondent knew before the training program, and the other is the number of the treated entrepreneurs with whom the respondent talked about training contents. The former is predetermined. Still it is probably endogenous because if the respondent is more sociable, he or she is likely to have a greater number of entrepreneurs and also likely to have higher business performance as a result of his or her good relationship with transaction partners and workers. The latter is more obviously endogenous.

Despite the endogeneity, we included these contact variables in the regression to see whether they have close association with the outcome variables when the influences of other variables are controlled for. A major finding is that the contact variables have highly significant coefficients soon after the training program, but their coefficients become insignificant in the subsequent years or survey rounds. In column 1, the interaction of the number of contacts and the first follow-up survey dummy has a positive and highly significant coefficient. Yet, the coefficient on the interaction of the number of contacts and the second follow-up survey dummy is negative and highly significant while the coefficient on the second follow-up survey dummy is positive and highly significant. These results indicate that knowledge spilt over to acquaintances initially but later on to other entrepreneurs. Column 2 shows a similar pattern even though significance levels are lower. Similarly, in column 3, the interaction of the number of contacts and the 2010 year dummy has a positive and marginally significant coefficient.

Another major finding is that in column 3, the impact of training on value added (as captured by the coefficient on the interaction of Both training and the 2013 year

dummy) is estimated to be positive and marginally significant. This result seems to reinforce the results shown in Table 5 and the conclusion of this paper.

## References

- Barro, R., Lee, J.W., 2012. A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics*, 104, 184-198.
- Berge, L.I.O., Bjorvatn, K., Juniwaty, K.S., Tungodden, B., 2012. Business training in Tanzania: From research-driven experiment to local implementation. *Journal of African Economies*, 21(5), 808-827.
- Berge, L.I.O., Bjorvatn, K., Tungodden, B., 2014. Human and financial capital for microenterprise development: Evidence from a field and lab experiment. *Management Science* 61, 707-722.
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., Roberts, J., 2013. Does management matter? Evidence from India. *Quarterly Journal of Economics*, 128(1), 1-51.
- Bloom, N., Genakos, C., Sadun, R., van Reenen, J., 2012. Management practices across firms and countries. *Academy of Management Perspective*, 26(1), 12-33.
- Bloom, N., Schankerman, M., van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. *Econometrica*, 81, 1347-1393.
- Bloom, N., Van Reenen, J., 2010. Why do management practices differ across firms and countries? *Journal of Economic Perspective*, 24(1), 203-224.
- Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N., Freeman, P., 2008. Eliciting willingness to pay without bias: Evidence from a field experiment. *Economic Journal*, 118(525), 114-137.
- Bruhn, M., Karlan, D., Schoar, A., 2010. What capital is missing in developing countries? *American Economic Review: Papers and Proceedings*, 100(2), 629-633.
- Bruhn, M., Zia, B., 2013. Stimulating managerial capital in emerging markets: The impact of business training for young entrepreneurs. *Journal of Development*

*Effectiveness*, 5(2), 232-266.

- Burbidge, J., Magee, L., Robb, A.L., 1988. Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83, 123-127.
- Campos, F., Coville, A., Fernandes, A.M., Goldstein, M., McKenzie, D., 2014. Learning from the experiments that never happened: Lessons from trying to conduct randomized evaluations of matching grant programs in Africa. *Journal of The Japanese and International Economies*, 30, 4-24.
- De Grip, A., Sauermann, J., 2012. The effects of training on own and co-worker productivity: Evidence from a field experiment. *Economic Journal*, 122, 376-399.
- De Mel, S., McKenzie, D., Woodruff, C., 2014. Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. *Journal of Development Economics*, 106, 199-210.
- De Mel, S., McKenzie, D., Woodruff, C., 2009. Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics*, 88, 19-32.
- Drexler, A., Fischer, G., Schoar, A., 2014. Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2), 1-31.
- Evans, D.S., 1987. Tests of alternative theories of firm growth. *Journal of Political Economy*, 95 (4), 657-676
- Field, E., Jayachandran, S., Pande, R., 2010. Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India. *American Economic Review*, 100 (2), 125-129.
- Higuchi, Y., Nam, V.H., and Sonobe, T. 2015. Sustained impacts of *Kaizen* training.

- Journal of Economic Behavior and Organization*, 120, 189-206.
- Imai, M., 2012. *Gemba Kaizen: A Commonsense Approach to Continuous Improvement Strategy*, Second Edition. New York: McGraw-hill.
- Imbens, G., Angrist, J., 1994. Identification and estimation of local average treatment effects. *Econometrica*, 62, 467-476.
- Karlan, D., Knight, R., Udry, C., 2012. Hoping to win, expected to lose: Theory and lessons on micro enterprise development. *NBER Working Paper No. 18325*.
- Karlan, D., Valdivia, M., 2011. Teaching entrepreneurship: Impact of business training of microfinance clients and institutions. *Review of Economics and Statistics*, 93(2), 510-527.
- Mano, Y., Iddrisu, A., Yoshino, Y., Sonobe, T., 2012. How can micro and small enterprises in Sub-Saharan Africa become more productive? The impacts of experimental basic management training. *World Development*, 40 (3), 458-68.
- McKenzie, D., 2012. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2), 210-221.
- McKenzie, D., Woodruff, C., 2015. Business practices in small firms in developing countries. *NBER Working Paper Series 21505*.
- McKenzie, D., Woodruff, C., 2014. What are we learning from business training and entrepreneurship evaluations around the developing world. *World Bank Research Observer*, 29(1), 48-82.
- Morgan, J., Liker, J.K., 2006. *The Toyota Product Development System: Integrating People, Process, and Technology*. London: Productivity Press.
- Rubin, D.B., 1978. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 6(1), 34-58.



Söderbom, M., Teal, F., Harding, A., 2006. The determinants of survival among African manufacturing firms. *Economic Development and Cultural Change*, 54 (3), 533-55.

**Table 1: Balance Check**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Group TT mean (s.d.)	Group TC mean (s.d.)	Group CT mean (s.d.)	Group CC mean (s.d.)	TT – CC <i>p</i> -value	TC – CC <i>p</i> -value	CT – CC <i>p</i> -value
Age	44.5	44.8	45.2	44.8	0.89	0.99	0.84
(as of the baseline survey)	(9.06)	(7.52)	(9.49)	(7.53)			
Male	0.08	0.17	0.14	0.24	0.10	0.51	0.36
(yes = 1)	(0.27)	(0.38)	(0.36)	(0.44)			
Chagga	0.15	0.25	0.32	0.31	0.18	0.64	0.93
(yes = 1)	(0.37)	(0.44)	(0.48)	(0.47)			
Years of education	11.3	10.3	10.6	10.7	0.44	0.62	0.91
	(2.62)	(2.12)	(2.66)	(2.90)			
Past business training experience	0.73	0.67	0.61	0.55	0.17	0.40	0.68
(yes = 1)	(0.45)	(0.48)	(0.50)	(0.51)			
Former textile employee	0.15	0.25	0.25	0.17	0.86	0.50	0.48
(yes = 1)	(0.37)	(0.44)	(0.44)	(0.38)			
Years of operation	11.9	11.8	12.0	10.5	0.36	0.40	0.35
(as of baseline survey)	(5.45)	(4.85)	(6.34)	(6.10)			
Member of TANCRAFT	0.42	0.46	0.46	0.55	0.35	0.51	0.52
(yes = 1)	(0.509)	(0.51)	(0.51)	(0.51)			
Member of HOT	0.46	0.46	0.54	0.31	0.25	0.28	0.09
(yes = 1)	(0.51)	(0.51)	(0.51)	(0.47)			
Member of ADAT	0.12	0.08	0.00	0.14	0.81	0.54	0.04
(yes = 1)	(0.33)	(0.28)	(0.00)	(0.35)			
No. enterprises in the Group	26	24	28	29			
No. participated in classroom training	22	23	0	0			
No. participated in on-site training	26	0	28	0			

Notes: Numbers in parenthesis are standard deviations. *P*-values are from the *t*-test concerning the null hypothesis that the mean values are the same in the two groups.

**Table 2: Changes in Management Score and Business Performance by Group**

	(1) Group TT	(2) Group TC	(3) Group CT	(4) Group CC	(5) TT – CC	(6) TC – CC	(7) CT – CC
<b>Management score, 0 to 27</b>							
Baseline	11.7	10.3	10.2	9.7	*	-	-
(March 2010)	[11]	[10]	[9.5]	[9]			
1st follow-up	17.7	16.8	14.5	12.4	***	***	-
(soon after training)	[17.5]	[15.5]	[14.5]	[11]			
2nd follow-up	20.3	19.7	19.8	17.0	***	***	***
(1.5 years after)	[22]	[20]	[19]	[18]			
3rd follow-up	16.4	16.9	15.6	12.6	***	***	***
(3 years after)	[17]	[16.5]	[15]	[12]			
<b>Value added [USD]</b>							
Baseline	14,473	13,551	12,895	12,838	-	-	-
(average of 2008&09)	[11,651]	[9,241]	[8,030]	[7,588]			
2010	24,623	20,750	14,789	24,364	-	-	-
	[20,248]	[15,730]	[9,432]	[18,993]			
2011	18,092	23,667	20,909	22,605	-	-	-
	[14,801]	[13,549]	[14,718]	[14,669]			
2012	17,380	12,059	16,445	12,574	-	-	-
	[12,544]	[9,106]	[9,069]	[9,087]			
2013	18,914	12,460	13,275	12,535	**	-	-
	[13,953]	[9,390]	[7,230]	[7,316]			
<b>Sales revenue [USD]</b>							
Baseline	23,328	23,130	21,235	19,726	-	-	-
(average of 2008&09)	[17,807]	[14,225]	[14,396]	[14,022]			
2010	34,527	33,337	24,204	39,484	-	-	-
	[29,881]	[25,101]	[17,623]	[25,101]			
2011	33,459	35,727	40,579	38,641	-	-	-
	[29,460]	[27,385]	[36,260]	[24,881]			
2012	32,379	24,453	31,515	25,132	-	-	-
	[26,390]	[19,353]	[21,112]	[17,074]			
2013	35,786	24,528	26,654	25,818	**	-	-
	[28,191]	[21,143]	[18,241]	[19,899]			
<b>Investment (yes = 1)</b>							
Jan. 2011- Sep. 2012	0.23	0.17	0.27	0.27	-	-	-
	(0.43)	(0.39)	(0.45)	(0.45)			
Jan. 2012- Dec. 2013	0.36	0.33	0.24	0.11	**	*	-
	(0.49)	(0.48)	(0.44)	(0.32)			

Notes: Numbers in bracket are medians while numbers in parenthesis are standard deviations. Columns 5 to 7 show the significance of the difference based on the *t*-test with asterisks. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% levels, respectively, and - indicates insignificance. Sales revenue and value added are in terms of international dollar converted by using the PPP conversion factor, GDP (LCU per international dollar), available at World Bank's *World Development Indicators*.

**Table 3: Knowledge Spillovers**

	(1) Group TT	(2) Group TC	(3) Group CT	(4) Group CC	(5) Total
(a) No. of sample entrepreneurs with whom the respondent has talked about training contents	12.2 (9.7)	11.1 (6.2)	6.6 (7.5)	3.7 (4.5)	8.2 (7.9)
(b) No. of training participants with whom the respondent has talked about training contents	11.0 (8.6)	9.9 (5.4)	5.5 (6.4)	3.4 (4.1)	7.3 (7.0)
(c) Willingness to pay (yes = 1)					
(c.1) Baseline	0.65 (0.49)	0.75 (0.44)	0.64 (0.49)	0.65 (0.49)	0.67 (0.47)
(c.2) 1st follow-up	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
No. enterprises in the Group	26	24	28	29	107

Notes: Numbers in parenthesis are standard deviations.

**Table 4: Impact on Management Score (ANCOVA)**

	(1)	(2)	(3)	(4)	(5)	(6)
	<u>Total Management</u>		<u>Sales</u>	<u>Record</u>	<u>Marketing</u>	<u>Kaizen</u>
	<u>Score</u>		<u>Promotion</u>	<u>Keeping</u>		
	ITT	TOT	TOT	TOT	TOT	TOT
Both training × 1st follow-up	5.00*** (5.55)	5.90*** (5.89)	0.87** (2.20)	1.01*** (3.86)	1.67*** (4.89)	2.35*** (3.86)
Both training × 2nd follow-up	1.69** (2.03)	2.08** (2.12)	0.51 (1.28)	0.31 (1.42)	0.22 (1.09)	1.04* (1.70)
Both training × 3rd follow-up	3.09*** (3.75)	3.71*** (3.79)	0.68** (2.09)	1.28*** (3.60)	0.19 (0.61)	1.55** (2.31)
Either training × 1st follow-up	2.68*** (3.17)	2.74*** (3.28)	0.50* (1.93)	0.38 (1.42)	0.38 (1.34)	1.49*** (3.04)
Either training × 2nd follow-up	1.78** (2.59)	1.81*** (2.70)	0.26 (1.01)	0.27 (1.51)	0.23 (1.51)	1.06** (2.39)
Either training × 3rd follow-up	3.39*** (4.30)	3.48*** (4.39)	0.77*** (2.96)	1.07*** (3.93)	0.41** (2.27)	1.23** (2.22)
Time dummy (2nd follow-up)	4.03*** (3.96)	4.05*** (4.07)	0.90*** (3.67)	0.62*** (3.49)	0.97*** (3.63)	1.56** (2.26)
Time dummy (3rd follow-up)	1.34 (1.39)	1.29 (1.35)	-0.56** (-2.13)	0.10 (0.41)	1.46*** (4.51)	0.29 (0.45)
First-stage <i>F</i> -statistics		147.6	147.6	147.6	147.6	147.6
Baseline mean of <i>y</i>	10.6	10.6	1.1	2.6	1.4	5.4

Notes: In all columns, the number of observations is 312, and the number of enterprises in the sample is 107. Numbers in parentheses are *t*-statistics based on standard errors clustered at the enterprise level. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent levels, respectively. In column (1), Both training is a dummy variable equal to 1 if the enterprise was assigned to Group TT and 0 otherwise, and Either training is a dummy variable equal to 1 if the enterprise is assigned to either Group TC or CT and 0 otherwise. In the other columns, Both training is a dummy variable equal to 1 if the enterprise participated in both components and 0 otherwise, and Either training is a dummy variable equal to 1 if the enterprise participated either the classroom or on-site component but not both, and 0 otherwise. All estimations include the baseline values of each dependent variable and the enumerator fixed effects as explanatory variables even though their estimated coefficients are not reported.

**Table 5: Impact on Business Performance (ANCOVA)**

	(1)	(2)	(3)	(4)	(5)	(6)
	<u>Value added</u>			<u>Sales revenue</u>		
	USD		transformed	USD		Transformed
	ITT	TOT	TOT	ITT	TOT	TOT
Both training × 2010	4729.8 (1.34)	5541.9 (1.37)	0.30 (1.40)	2065.2 (0.34)	2408.9 (0.35)	0.19 (0.99)
Both training × 2011	-1361.7 (-0.33)	-1617.8 (-0.34)	0.031 (0.11)	-219.7 (-0.03)	-260.4 (-0.04)	0.13 (0.56)
Both training × 2012	4057.7 (1.15)	4905.0 (1.16)	0.44 (1.64)	5761.2 (1.06)	6961.9 (1.09)	0.28 (1.12)
Both training × 2013	6315.0** (2.20)	7512.2** (2.18)	0.47* (1.69)	9815.9* (1.80)	11673.8* (1.83)	0.31 (1.25)
Either training × 2010	-622.1 (-0.24)	-625.7 (-0.24)	-0.014 (-0.08)	-2253.1 (-0.39)	-2296.2 (-0.40)	0.016 (0.10)
Either training × 2011	2870.1 (0.74)	2883.7 (0.76)	0.15 (0.70)	5044.6 (0.81)	5050.2 (0.84)	0.19 (1.03)
Either training × 2012	-640.9 (-0.23)	-644.6 (-0.23)	0.16 (0.83)	-1779.8 (-0.37)	-1809.7 (-0.38)	0.063 (0.40)
Either training × 2013	1050.5 (0.43)	1080.7 (0.45)	0.18 (0.93)	533.1 (0.12)	550.8 (0.12)	0.10 (0.62)
Time dummy (2011)	2895.8 (0.94)	2967.4 (0.99)	0.11 (0.61)	4269.7 (0.99)	4300.2 (1.02)	0.14 (0.96)
Time dummy (2012)	-2340.9 (-0.80)	-2414.7 (-0.85)	-0.35* (-1.74)	-4034.4 (-0.69)	-4123.4 (-0.73)	-0.14 (-0.80)
Time dummy (2013)	-2961.4 (-0.94)	-3035.2 (-0.98)	-0.28 (-1.26)	-4562.0 (-0.71)	-4651.1 (-0.74)	-0.10 (-0.55)
1st-stage <i>F</i> -statistics		135.7	135.5		135.7	135.5
Baseline mean of <i>y</i>	13609	13609	9.06	21760	21760	9.58

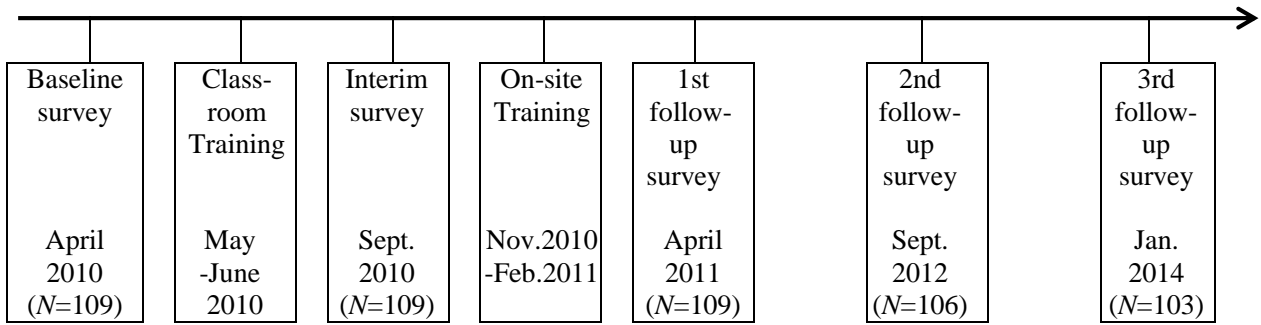
Notes: In all columns, the number of observations is 410, and the number of enterprises in the sample is 107. Numbers in parentheses are *t*-statistics based on standard errors clustered at the enterprise level. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent levels, respectively. In column (1), Both training is a dummy variable equal to 1 if the enterprise was assigned to Group TT and 0 otherwise, and Either training is a dummy variable equal to 1 if the enterprise is assigned to either Group TC or CT and 0 otherwise. In the other columns, Both training is a dummy variable equal to 1 if the enterprise participated in both components and 0 otherwise, and Either training is a dummy variable equal to 1 if the enterprise participated either the classroom or on-site component but not both, and 0 otherwise. All estimations include the baseline values of each dependent variable and the enumerator fixed effects as explanatory variables even though their estimated coefficients are not reported.

**Table 6: ITT Impact of the Training on Capital Investment, Cross Section**

	(1)	(2)	(3)	(4)
	<u>Jan. 2011-Sep. 2012</u>		<u>Jan. 2012-Dec. 2013</u>	
	y = 1 if made any investment Probit model, Marginal effects	Amount of investment OLS, Coefficients	y = 1 if made any investment Probit model, Marginal effects	Amount of investment OLS, Coefficients
Both training	-0.014 (-0.12)	-160.4 (-0.41)	0.30** (2.07)	177.6 (0.12)
Either training	-0.052 (-0.53)	490.8 (0.81)	0.20* (1.65)	689.8 (0.39)
No. observations	103	103	101	101

Notes: Numbers in parentheses are *t*-statistics based on standard errors robust to heteroscedasticity. \*\*, and \* indicate significance at the 5 and 10 percent levels, respectively. Both training is a dummy variable equal to 1 if the enterprise was assigned to Group TT and 0 otherwise, and Either training is a dummy variable equal to 1 if the enterprise is assigned to either Group TC or CT and 0 otherwise. All estimations include the baseline values of each dependent variable and the enumerator fixed effects as explanatory variables, but their coefficients are not reported.

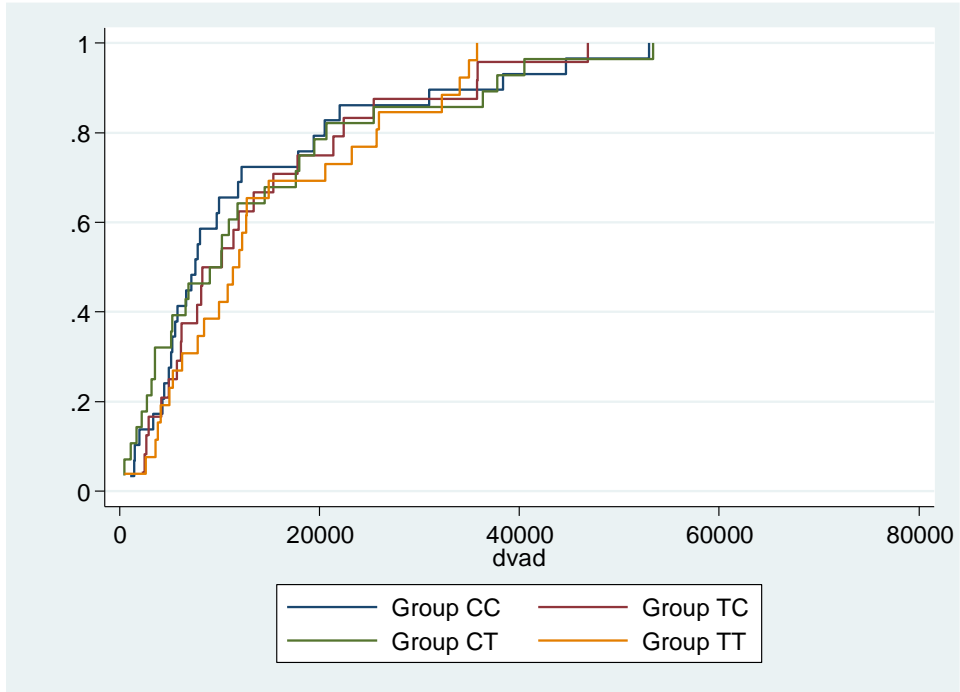
**Figure 1: Timeline**



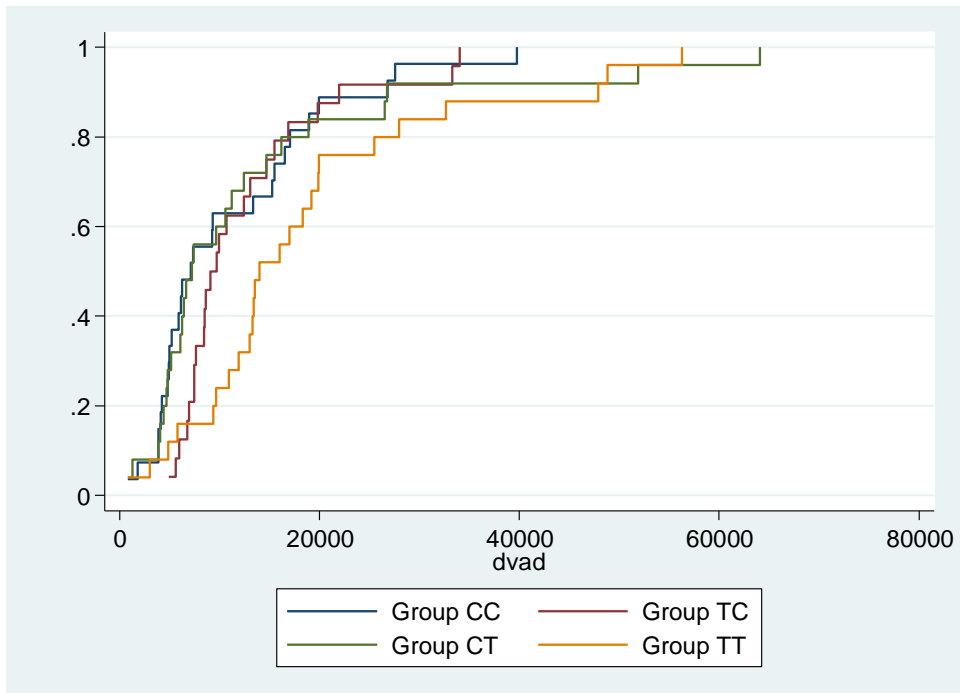


**Figure 2: CDF of annual value added by group**

(a) At baseline survey (2009)



(b) At fourth follow-up (2013)



**Appendix Table A1:  
Management Score Components and Baseline Adoption Rates (%)**

---

<i>Sales promotion</i>	
The enterprise has had any expenditure for advertisement in the last 3 months.	10
The enterprise has any signboards in front of the workshop.	39
The enterprise distributes complimentary cards or calendar?	27
The enterprise issues invoices or receipts with its name or phone number.	36
<i>Record keeping</i>	
The enterprise preserves records (e.g., receipts or invoices) of purchase.	48
The enterprise separates business and household expenses.	62
The enterprise keeps record of sales.	84
The enterprise keeps record of material purchase.	70
<i>Marketing</i>	
The entrepreneur can clearly describe the characteristics of their customers.	42
The entrepreneur can clearly describe its strength vis-a-vis its competitors.	24
The entrepreneur has clear sales or profit target in this year.	45
The entrepreneur has clear plan for its growth in the next five years.	28
<i>Kaizen</i>	
The enterprise assigns any workers to inspect product quality before shipping.	10
The enterprise keeps records of quality defects.	22
The enterprise records customers' complaints about the products sold.	45
The enterprise instructs the worker the way of preventing the defect.	9
The enterprise has a designated place for all tools.	34
The enterprise labels tool storage so that workers can easily find necessary tools.	3
The enterprise has a designated place for raw material storage.	76
The enterprise separately stores raw materials from the scrap.	75
The enterprise has no scrap cloths around the floor.	13
The enterprise daily removes scraps and cleans the floor of the workplace.	83
The enterprise does machine maintenance at least once a week.	29
The enterprise holds regular meetings of all production workers.	28
The enterprise has a designated area for all production activities in its workshop.	29
The enterprise posts a flow chart describing the proper production process.	8
The entrepreneur is aware of the sequence and duration of each production process.	82

---

**Appendix Table A2:  
Baseline Correlates of Management Score and Business Performance (OLS)**

	(1) <u>Baseline management score</u>	(2) <u>Baseline value added</u>	(3)	(4) <u>Baseline Sales revenue</u>	(5)
Age	-0.55 (-0.99)	-147.2 (-0.09)	357.2 (0.27)	-734.9 (-0.33)	44.1 (0.02)
Age squared	0.0049 (0.82)	1.63 (0.09)	-2.83 (-0.20)	7.65 (0.33)	0.77 (0.04)
Male (yes = 1)	-1.41* (-1.68)	281.1 (0.09)	1566.0 (0.54)	2618.0 (0.51)	4602.5 (0.92)
Chagga (yes = 1)	0.68 (0.96)	5520.1* (1.85)	4901.8 (1.66)	7853.3* (1.90)	6898.3* (1.72)
Years of education	0.31*** (2.65)	987.6** (2.28)	705.7* (1.73)	1349.1** (2.17)	913.8 (1.56)
Past business training experience (yes = 1)	1.17 (1.56)	4291.8* (1.76)	3224.8 (1.34)	7850.2** (2.39)	6202.4* (1.94)
Former textile employee (yes = 1)	-0.99 (-1.01)	-1337.9 (-0.52)	-437.0 (-0.18)	-1805.2 (-0.43)	-413.9 (-0.10)
Years of operation	0.13** (2.34)	640.1*** (2.75)	521.8** (2.30)	1109.3*** (2.97)	926.4** (2.62)
Baseline management score			910.3** (2.33)		1405.9*** (2.72)
No. observations	107	107	107	107	107
R <sup>2</sup>	0.276	0.280	0.324	0.330	0.379
Adjusted R <sup>2</sup>	0.209	0.214	0.254	0.268	0.314

Notes: The sample size is 107. Numbers in parentheses are *t*-statistics based on standard errors robust to heteroscedasticity. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent levels, respectively. All estimations include the enumerator fixed effects as explanatory variables, but their estimates are not reported.

**Appendix Table A3: Spillover (ANCOVA, TOT)**

	(1)	(2)	(3)	(4)	(5)	(6)
	<u>Management score</u>		<u>Value added</u>		<u>Sales revenue</u>	
	known	talked	known	talked	known	talked
Both training			2566.3	2853.2	-2186.9	-1053.9
× 2010			(0.67)	(0.73)	(-0.29)	(-0.14)
Both training	5.19***	5.34***	-5887.9	-6822.4	-5876.6	-7083.7
× 1st follow-up or 2011	(4.99)	(4.79)	(-1.24)	(-1.45)	(-0.75)	(-0.89)
Both training	1.70*	1.38	1967.7	1018.4	2181.7	967.3
× 2nd follow-up or 2012	(1.83)	(1.43)	(0.48)	(0.24)	(0.35)	(0.15)
Both training	2.98***	2.66***	5561.2*	4611.5	7856.7	6638.5
× 3rd follow-up or 2013	(2.99)	(2.66)	(1.66)	(1.32)	(1.21)	(1.01)
Either training			-2981.0	-1974.9	-6156.0	-4174.9
× 2010			(-1.02)	(-0.73)	(-0.94)	(-0.69)
Either training	2.10**	2.50***	-197.9	-353.8	1135.4	937.2
× 1st follow-up or 2011	(2.40)	(2.81)	(-0.05)	(-0.09)	(0.18)	(0.15)
Either training	1.40**	1.36**	-2255.3	-2380.6	-4264.8	-4420.5
× 2nd follow-up or 2012	(2.27)	(2.16)	(-0.69)	(-0.73)	(-0.82)	(-0.85)
Either training	2.69***	2.67***	-883.4	-1002.3	-2595.5	-2741.0
× 3rd follow-up or 2013	(3.45)	(3.43)	(-0.33)	(-0.38)	(-0.52)	(-0.55)
No. contacts			241.2**	281.7*	383.0**	362.9
× 2010			(2.43)	(1.68)	(2.28)	(1.28)
No. contacts	0.085***	0.089*	58.3	182.9	0.65	231.0
× 1st follow-up or 2011	(3.97)	(1.75)	(0.41)	(1.26)	(0.00)	(1.43)
No. contacts	-0.072***	-0.025	-45.1	77.3	-60.2	167.6
× 2nd follow-up or 2012	(-3.05)	(-1.07)	(-0.37)	(0.73)	(-0.35)	(1.15)
No. contacts	-0.024	0.021	-103.5	17.8	-112.5	114.0
× 3rd follow-up or 2013	(-0.92)	(0.85)	(-0.91)	(0.18)	(-0.58)	(0.71)
Time dummy			1456.2	-519.0	3635.2	57.1
(2011)			(0.40)	(-0.14)	(0.73)	(0.01)
Time dummy	5.39***	4.67***	-1493.2	-3432.8	-2533.6	-6029.4
(2nd follow-up or 2012)	(5.21)	(4.47)	(-0.45)	(-1.10)	(-0.44)	(-1.02)
Time dummy	1.79*	1.07	-1216.3	-3138.7	-2261.5	-5735.8
(3 years after or 2013)	(1.81)	(1.07)	(-0.35)	(-0.91)	(-0.38)	(-0.93)
No. observations	299	299	393	393	393	393
No. enterprises	101	101	101	101	101	101

Notes: Numbers in parentheses are *t*-statistics based on standard errors clustered at the enterprise level. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent levels, respectively. Both training is a dummy variable equal to 1 if the enterprise participated in both components and 0 otherwise, and Either training is a dummy variable equal to 1 if the enterprise participated either the classroom or on-site component but not both, and 0 otherwise.