Short- and Medium-Run Impacts of Management Training: An Experiment in Tanzania

By YUKI HIGUCHI, EDWIN P. MHEDE, AND TETSUSHI SONOBE*

We conducted a randomized controlled trial of management training in Tanzania, collecting baseline and follow-up data over the span of four years with negligible incidence of attrition. The data reveal that the training intervention improved the management practices of the treated firms, and that the impact remained significant three years after the intervention. Moreover, the initially insignificant impacts on value added became significant later, particularly in a sub-group of the treated firms that received both the classroom training and on-site consultation. These firms made adaptive efforts to select useful practices and to modify them to fit their business operation. (*JEL* L2, M1, O1)

* Higuchi: Graduate School of Economics, Nagoya City University, 1 Yamanohata, Mizuho-cho, Mizuho-ku, Nagoya, 467-0802, Japan (email: higuchi@econ.nagoya-cu.ac.jp); Mhede: Ministry of Industry, Trade, and Investment, NSSF Waterfront House, Gerezani Street, Dar es Salaam, Tanzania (email: meptz@yahoo.com); Sonobe: National Graduate Institute for Policy Studies, 7-22-1, Roppongi, Minato-ku, Tokyo, 106-8677, Japan (email: sonobete@grips.ac.jp). We thank Paulo Bastos, Nicholas Bloom, Takashi Kurosaki, David McKenzie, Keijiro Otsuka, Armando J. Garcia Pires, Yasuyuki Todo, Shinsuke Uchida, and John van Reenen for helpful comments. 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 many management practices soon after the training.
- Later they focused on and assimilated a smaller number of useful practices.

Keywords
Africa, Randomized controlled trial, Management training, *Kaizen*, Medium-run impacts

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 Bloom, Genakos, et al. (2012) and McKenzie and Woodruff (2017) among others attest. There has been renewed interest in managerial capacity building through business training in the recent years. An emerging body of literature finds, on the basis of randomized controlled trials (RCTs), that training can improve management practices (e.g., Anderson-Macdonald 2014; Berge, Bjorvatn, Juniwaty et al. 2012; Berge, Bjorvatn, and Tungodden 2014; De Mel, McKenzie, and Woodruff 2014; Drexler, Fischer, and Schoar 2014; 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. This is because 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 sales revenue, value added, or productivity. This 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 to see impacts on financial indicators compared to impacts on attitudes and practices (e.g., Imai 2012; Morgan and Liker 2006). This possibility seems to have been explored only by Bruhn, Karlan, and Schoar (2017), Bloom, Eifert et al. (2013), and Higuchi, Nam, and Sonobe (2015) so far.

The present paper investigates what happens if the period of follow-up observation is extended. Like Bloom, Eifert et al. (2013) and unlike Bruhn et al. (2017),

we examine the impacts of training programs that feature production management and quality control. The subjects of our experiments operate much smaller businesses than the subjects of the Bloom, Eifert et al. (2013) experiment. They design, cut, and sew garment products in Dar es Salaam, Tanzania. We chose garment producers as experiment subjects because garment production is as ubiquitous in the developing world as spinning and weaving. We focus on garment producers located in the same city and facing the same set of markets for materials, labor, capital, and products in order to mitigate the difficulties arising from unobservable heterogeneity among enterprises. Higuchi et al. (2015) report findings from follow-up observation for two years in Vietnam, whereas this paper is based on three-year observations in Tanzania.

We conducted a baseline survey of randomly selected garment manufacturers in early 2010 and then assigned them randomly to treatment and control groups. The treated enterprises were invited to participate in a classroom training program, an on-site training program, or both programs. Classroom training participants learned from trainers about good management practices for nearly 40 hours in total. On-site training participants received concrete advice on how to improve efficiency and safety at work from the trainers who visited their workshops several times. Follow-up surveys were conducted in 2011, 2012, and 2014 to collect data on management practices and the values of sales revenues and costs. 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 treatment groups (i.e. those invited to one of the three training programs) applied a significantly greater number of good management practices soon after the training than the control group. 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 two additional major findings. First, while the treatment groups adopted a number of good practices at first, they later stopped using some of these practices because they did not find them to be very useful. Second, participation in the combined classroom and on-site training program was found to have increasingly strong impacts on annual value added and sales revenue over time, and 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 taught. A possible explanation was given by some training participants, who told us that while they initially put into practice as many training lessons as possible, they gradually discarded the ones that were not found to be 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 and familiar with the Start/Improve Your Business (SIYB) training materials developed by the International Labor Organization and widely used in developing countries.[1] Our training programs, 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 subjects are small enterprises, they are somewhat larger in employment size than microenterprises and microcredit clients, the typical beneficiaries of the above type of business training. For this reason, our program

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

covered know-how for coordinating the division of labor among workers, featuring basic *Kaizen*, a common-sense approach to production management and quality control.

Presumably because the sample enterprises were not very small in operation size, sample attrition due to exit from market is not serious in this study. Because the sample enterprises were drawn from a single industry in the same city, the heterogeneity to be controlled must be relatively small in this study. Information spillovers are beyond control, however. Many training participants were so satisfied with the training program that they could not keep what they learned at the program secret to their friends including those in the control group. Data on spillovers are presented below just as supplementary information because of endogeneity: those participants who anticipated positive training impacts tended to talk to acquaintances about the training content. Our impact assessment assumes away spillovers. Hence the assessment results may understate the training impacts.

The remainder of the paper is organized as follows. Section I puts forward our hypotheses and describes the experimental design. Section II addresses the issues of balance, attrition, and externality, while Section III presents the impact evaluation results. Finally, Section IV summarizes the findings and discusses implications for future studies.

I. Hypotheses and Experimental Design

A. Study Site and Sample Enterprises

In Dar es Salaam, there are about 250 small-scale firms and numerous self-employed businesses producing garment products. The majority of these small-scale firms are owned and operated by female entrepreneurs. Presumably this is because of a sewing

business training program provided by the United Nations Industrial Development Organization (UNIDO) in the 1990s to nurture female entrepreneurs. Some participants' sewing businesses became successful, which encouraged many other females and a smaller number of males in the city to start their own sewing businesses. In this way, a cluster was formed.

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).[2] Members of these associations sew dresses, shirts, jackets, and textile homeware, such as cushion covers and kitchen mittens.[3] Out of these members, we selected 113 sample enterprises randomly for our survey. They are located in the same city, faced with the same or related markets for products and inputs, and share the same information on market and technology which they receive from the associations.

B. Hypotheses and Randomization

Many studies of management training RCTs have found that the management practices that trainers have taught to trainees come into use quickly. But the existing evidence for favorable impacts on business performance is relatively weak (McKenzie and Woodruff 2014). Our basic hypothesis is that this is because it takes managers and workers some time to assimilate new practices. It seems possible that attempts at using a new practice

2 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 the profits.

3 For many members, the associations are useful because they provide information on various training programs and events and because they organize trade fairs. Interested members participate in trade fairs held in neighboring countries as well as in Tanzania once or twice a year. They charter buses, carry sacks of their products with them, and enjoy conversations with colleagues and travel.

lower productivity temporarily before managers and workers get accustomed to them. Moreover, business performance is subject to a variety of shocks to enterprises. Among those shocks, some shocks, such as fluctuations in product market demand and input prices, could be controlled for in regression analysis if sample firms are relatively homogeneous like ours. However, there are unobservable idiosyncratic shocks as well, including illness, fires, contract breach, and so on. Thus, we advance the following testable hypothesis:

Hypothesis 1: A significant effect of a management training program on business performance is detected later than that on management practices.

In addition to time to assimilation and idiosyncratic shocks, various factors may affect the magnitude and time lag of training impacts. The intensity, quantity, and quality of training may be among those factors. In principle, an experimenter could assign experiment subjects randomly to different training programs with different intensity, quantity, or quality. In reality, however, it seems difficult to vary training intensity, quality, or length substantially because business owners may stop cooperating with the experiment if training is too intense, if they find themselves receiving low-quality training and hence wasting time, or if a long program is just lengthy and boring.

As mentioned earlier and will be explained in detail below, we prepared two short programs called classroom and on-site training programs. In the classroom program, participants would gather in a classroom to receive lectures from trainers. In the on-site program, trainers would visit and observe participants' workshops several times to give

advice. The classroom program would be held earlier than the on-site program so as to allow some sample enterprises to participate in both programs. This combined program would serve as a long program. We assigned the 113 sample enterprises randomly to a control group, a class-only group (i.e., those enterprises which were invited to only the classroom program), a onsite-only group (i.e., those to only the on-site program), and a class+onsite group (i.e., those to the combined program). Comparing expected impacts of the short and combined programs, we hypothesize as follows:

Hypothesis 2: Compared with two short training programs, the combined program has relatively strong and quick effects on management practices and business performance.

C. Training Content and Timeline

Most of the existing studies of management training interventions highlight the SIYB type of basic training in accounting, marketing, and business strategy (McKenzie and Woodruff 2014). Such training is suitable to microenterprises with zero to a few employees. By contrast, our sample enterprises employed a little more than five workers on average. Thus, the training intervention in this study included, in addition to SIYB training, the introductory lessons of production management and quality control. Coordinating the division of labor among workers, reducing wasteful use of materials and time, and preventive maintenance were among major topics in the training program. *Kaizen* is widely used in both developed and developing countries. It is a Japan-pioneered approach based on the US-born industrial engineering and statistical quality control, and is the origin of Lean Manufacturing and Six Sigma. It emphasizes productivity improvement by the collaborative and continuous efforts of both managers

and workers (Imai 2012; Hosono 2017).[4] Our training program featured the elementary 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 experiences 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, had not been very familiar with *Kaizen*. Thus, we made arrangements 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 the Japanese expert taught the training participants directly as well, the local trainers handled most of the classroom and on-site programs.

The timeline of the experiment is shown in Figure 1. We conducted a baseline survey of enterprises in April 2010, provided the classroom training program in May and June 2010, and conducted a quick interim survey, which was followed by the on-site training program and three full-scale follow-up surveys. The baseline survey collected detailed and reliable data from 113 enterprises. The classroom training program took 2.5 hours a day, five days a week over a three-week span to teach the basic knowledge of *Kaizen* and the standard content of the SYIB training in entrepreneurship, marketing, and record keeping.[5] To the classroom program, we invited the owners or top managers (hereafter referred to as entrepreneurs) of 52 randomly selected enterprises, of which 47 attended the training for at least 10 out of the

4 In the 1980s, the US automobile industry adapted *Kaizen* to their circumstances and named their version lean manufacturing.

5 The classroom training was supplemented by a one-week optional course on color coordination and product design.

15 days (or three weeks). We regard these 47 entrepreneurs as participants in this training program.

[Insert Figure 1 Here]

During the period between the two short programs, the *Kaizen* expert and the local trainers chose two participants in the classroom program to turn their workshops into model workshops. The purpose of making model workshops was twofold. First, model workshops would serve as real examples of streamlined workshop layout, good maintenance of machinery, and good use of shelves and smaller small items to sort and store materials, intermediate inputs, and work in progress. Second, when model workshops were being made, they would serve as venues where the local trainers learned from the *Kaizen* expert about how to spot inefficiencies, how to find implementable solutions, and how to carry out relatively large-scale changes, such as layout change. The expert and local trainers selected two enterprises that had relatively spacious workshops and product lines common to many sample enterprises and that agreed to show their layouts and *Kaizen* practices to other training participants. Since these two model workshops received special treatment, they are excluded from our sample in the analysis below.

From among the 111 remaining enterprises, we selected 55 enterprises randomly to invite to the on-site training program. It began with a half-day seminar, which all the invited entrepreneurs attended. At the seminar, they were briefed about the on-site training and *Kaizen* management and were randomly assigned to trainers. Each trainer visited the workshops under his or her charge to assess the situations and develop 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 visits was determined by the local consultants, depending on the willingness and availability of the respective participants. The trainers explained to the participants how to introduce some *Kaizen* practices to their workers and recommended some participants to try layout change. If requested by the participants, the trainers also gave advice on record keeping and marketing.

[Insert Table 1 Here]

D. Take-up Rate

As shown toward the bottom of Table 1, there are 26 enterprises in the class+onsite group, 24 in the class-only group, 29 in the onsite-only group, and 32 in the control group. This classification is based on intention to treat, i.e., on invitations, as the result of randomization. Four enterprises in the class+onsite group and one enterprise in class-only group did not participate in the classroom training component. Thus, the take-up rate for the classroom training was 90 percent. The take-up rate for the on-site training was 100 percent. All the participants were those invited to the respective programs, and there were no training participants who were not invited.

A possible reason for the high take-up rate for the on-site program is that the participants did not have to travel to a training venue in the case of the on-site training. Another possible reason is that the sample entrepreneurs became more willing to receive training after the classroom training program had established a good reputation, as we will discuss shortly. There is a caveat, however. Some entrepreneurs in the onsite-only

group, i.e., those who had not participated in the classroom program, might fail to take the on-site training seriously and fail to keep their appointments with the trainers. Some did not seem to take the trainers into their confidence even after they received the first visit from the trainers. These participants in the on-site program would miss some useful suggestions.

II. Internal Validity

A. Balance

During the baseline and follow-up surveys, our enumerators visited the 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 the data on the characteristics of the sample enterprises and their entrepreneurs by treatment status (on the invitation, not participation, basis). The entrepreneurs are in their mid-40s on average, and the majority is female. Chagga, an economically active ethnic group, accounts for a quarter of our sample. The educational attainment of the 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) cross-country human capital data set. Compared with male entrepreneurs, female entrepreneurs are more highly educated. More than 60 percent of the whole sample and an even higher percentage of the female sub-sample had business training experience in the past. Only a small number of entrepreneurs had prior work experience in garment manufacturing.

The management practices of the sample enterprises were evaluated by our enumerators based on either visual inspection or the way that the entrepreneurs

responded to their questions or both. The enumerators used 27 diagnostic criteria, which are listed in Appendix Table 1. The evaluation results are expressed by the number of diagnostic criteria that the enumerator finds the enterprise satisfies. This number is henceforth called management score and ranges from zero to 27. As indicators of business performance, we use annual value added, sales revenue, employment size, and capital investment.[6] We refer to these performance indicators and the management score collectively as outcome variables. Appendix A gives the details of our baseline and follow-up surveys, the construction of the data, and our enumerators. Appendix B reports the correlation among the outcome variables and enterprise characteristics.

Columns 5 to 7 of Table 1 report the *p*-values for the tests of equality in the mean between each treatment group and the control group. Since the relatively small sample is subdivided into four groups, it is little wonder that there are two marginally significant differences at the 10 percent significance level. As shown toward the bottom of Table 1, however, the joint orthogonality test does not reject the null hypothesis that the treatment status is uncorrelated with any of the control and outcome variables. In this sense, randomization has successfully generated balance.

B. Attrition

There was no incidence of sample attrition in the first follow-up survey. When the second follow-up survey was conducted in September 2012, three sample enterprises (one in the class-only group and two in the onsite-only Group) were not in operation, and their data were not collected. While the third follow-up survey in January 2014,

6 The results of the impact evaluation reported in this paper remain qualitatively the same and quantitatively similar if value added and sales revenues are deflated by applying the GDP deflator.

however, found two of the three previously missing enterprises (one in the class-only group and one in the onsite-only group) had come back into operation, this survey could not collect data from six sample enterprises (one in the class+onsite group, three in the onsite-only group, and two in the control group) that were not in operation.

Figure 1 presents the numbers of enterprises interviewed in the surveys, including the two model workshops, even though the model workshops are excluded from the analysis below. The attrition rate, (113 – 107)/113, is low relative to the attrition rates reported by many other studies of management training experiments (e.g., McKenzie and Woodruff 2014). This may be attributed to the relatively large firm size of our sample enterprises because the studies of firm survival, such as Evans (1987) and Söderbom, Teal, and Harding (2006), suggest that smaller firms are more short-lived.

We do not find that survival is correlated with the treatment status. As shown in column 6 of Appendix Table 2, for example, whether or not a sample enterprise was operating during the third follow-up survey does not depend on its treatment status. Related to attrition, a question may arise as to why the control group continued to cooperate with us in the three follow-up surveys even though they were not invited to either component of the training program. A likely reason is that they expected to participate in our future training program.

C. Externality

All the sample enterprises are located in Dar es Salaam. Because of their geographical proximity and human network, knowledge could easily spill over from the training

participants to non-participants.[7] During the second follow-up survey, we showed the list of all the sample entrepreneurs, together with their enterprise names and addresses, to each sample entrepreneur and asked with whom on the list he or she had talked about the training content. Row (a) of Table 2 summarizes the answers to this question. For example, entrepreneurs in the class+onsite group or the class-only group talked with 12.2 or 11.6 other sample entrepreneurs on average, whereas the onsite-only training participants talked with only 7.6 other sample entrepreneurs. These results should not be surprising because for the onsite-only training participants, the kick-off meeting was the only opportunity for face-to-face contacts with other participants offered by the program. The control-group entrepreneurs talked with only 4 other entrepreneurs on average.

[Insert Table 2 Here]

Row (b) addresses the question of how much the sample entrepreneurs talked with the training participants about the training. An entrepreneur in the class+onsite group talked with 11 other training participants. In short, participants tended to talk with participants rather than non-participants. Although non-participants also shared the same tendency, they talked about the training contents much less than participants. Less direct evidence for knowledge spillovers includes the finding that the control group mean of the management score increased from 10.3 in the baseline survey to 12.8 in the first follow-up survey and further to 17.5 in the second follow-up survey (see column 4 in Table 1 and Panel A in Table 3).

7 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.

Suggestion of knowledge spillovers comes also from 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 combined program after explaining the purpose and features of the program.[8] Row (c) of Table 2 shows the fraction of the sample entrepreneurs who answered affirmatively. This proportion increased to 100 percent after the training program in all groups including the control group. The increased willingness to pay among the control group indicates that information on the training program spilt over from the training participants to them.

Since there seem to be knowledge spillovers, we attempted to take account of knowledge spillovers in assessing training impacts by using the conversation data. The results are reported in Appendix C. The use of the conversation data, however, does not necessarily improve the accuracy of impact evaluation because of the possibly endogenous nature of the contact data. Some participants might share new knowledge with friends because they were excited about the prospect that the use of the knowledge would increase profits. Those who talked less about the training content with others might be less convinced of increased profits. If this is the case, the use of the conversation data may invite an estimation bias, even though it may allow us to control the spillover effects to some extent.

The training impacts are estimated in the next section under the assumption that the control group is not at all affected by the training intervention, i.e., the stable unit

8 The question on willingness to pay is prone to reporting bias because of its hypothetical nature. 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?" If the answer was affirmative, we asked further, "How sure are you about the answer? Are you definitely sure or probably sure?" Based on this 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.

treatment value assumption (Rubin 1978). It follows that knowledge spillovers are assumed away, which in turn causes downward bias in the estimates of the training impacts. In general, however, the estimates of business training impacts based on the stable unit treatment value assumption can be overestimates. This is because the market supply of the treatment group in a training experiment may lower the sale revenues of their rivals including the control group through a lowed market price or through business stealing.

Our expectation is that while such market rivalry may exist in our study site, it will not cause a serious estimation bias for two reasons. First, as garment markets in East Africa are increasingly integrated, the majority of the sample enterprises export their products to neighboring countries in small quantities, and Dar es Salaam imports garment products in large quantities. The impact of the treatment groups' sales increases on the control group would be limited.[9] Second, our training program did not highlight sales promotion and marketing strategy as much as reductions in the wasteful use of intermediate inputs and time. As we will see below, a large part of training impact is expected to come from waste reductions rather than from increase in market supply. Thus, while both upward and downward biases are conceivable, we expect that the stable unit treatment value assumption leads to underestimates in our case.

III. Results

A. Econometric Specification

9 The situation seems similar to the one in the growing market in Kenya, where McKenzie and Puerto (2017) find that small enterprises, after receiving business training, improved consumer services and launched new products without causing negative effects on control group enterprises in the same market.

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

$$y_{it} = \alpha + \beta^{BOTH}{}_t Z^{BOTH}{}_i + \beta^{CLASS}{}_t Z^{CLASS}{}_i + \beta^{ONSITE}{}_t Z^{ONSITE}{}_i + \gamma y_{i0} + \sum_{n=1}^{N-1} \delta_n m_{nit} + \eta_t + \varepsilon_{it}, \quad (1)$$

where y_{it} is the management score or the employment size of enterprise i at the t-th round of the follow-up survey (t = 1, 2, 3) or the enterprise's value added or sales revenue in year t (t = 2010, 2011, 2012, 2013).[10] $Z^{BOTH}{}_i$ is a dummy variable indicating whether enterprise i was invited to both classroom and on-site training programs (i.e., whether the enterprise belongs to the class+onsite group) or not. Similarly, $Z^{CLASS}{}_i$ and $Z^{ONSITE}{}_i$ are dummy variables indicating whether the enterprise is in the classroom-only group or not and whether it is in the onsite-only group, respectively. Since we expect the training effects to change over time, the coefficients on these dummy variables, $\beta^{BOTH}{}_t$, $\beta^{CLASS}{}_t$, and $\beta^{ONSITE}{}_t$ have subscript t. In other words, three coefficients are estimated on each treatment group dummy if the dependent variable is the management score or the employment size and four coefficients if the dependent variable is the value added or the sales revenue. Taking advantage of the high compliance rate, we estimate these coefficients by the intention-to-treat (ITT) specification.

Following the lead of McKenzie (2012), we employ the ANCOVA estimator, which is more efficient than the fixed-effect model estimator. Specifically, the right-hand side of equation (1) includes the baseline value of the dependent variable, y_{i0}.

10 The outcome variable can also be a dummy variable taking the value of one if an enterprise made any investment in the period January 2011 - September 2012, or that in the period January 2012 - December 2013 (see Panel D in Table 4). Since the number of enterprises that made investments is not large and the amount of reported investment varies greatly, we use a dummy variable indicating whether or not investment was undertaken.

When y is value added or sales revenue, the baseline value is the mean of the values in 2008 and 2009 since the use of average baseline value improves efficiency, even though different specifications are also attempted as shown in Table 5 (McKenzie 2012).

Equation (1) 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. The purpose is to control for enumerator fixed effects, δ_n, i.e., heterogeneity in the way the enumerators evaluate management practices and elicit information. Since, in total, eight enumerators were engaged in the three follow-up surveys (i.e., $N = 8$), equation (1) has seven enumerator dummies.[11] The time effects common to all enterprises, η_t, are captured by two time dummies for the management score (or the employment size) and three time dummies for the value added (or the sales revenue). The error term, ε_{it}, may have autocorrelation within respective sample enterprises. We report hypothesis testing results based on standard errors clustered at the enterprise level.

B. Impacts on Management Scores

Table 3 shows the estimated impacts of the training on the management scores of the three treatment groups as compared with those of the control group. Panel A report the estimated impacts on the management score, whereas the other four panels report the estimated impacts on disaggregated scores.[12] The first three rows in each panel show the estimated coefficients $\beta^{BOTH}{}_t$, $\beta^{CLASS}{}_t$, and $\beta^{ONSITE}{}_t$ for each data point. They also show the p-values for two null hypotheses: one is that the coefficient is constant over time, and the other is that the coefficient is always zero. The fourth row in each panel

11 Although the estimated coefficients on the enumerator dummies are not reported, the null hypothesis that the coefficients are all zero is rejected in most cases, indicating that the inclusion of these dummy variables reduces estimation bias.

12 The way of disaggregation should be clear from the list of practices shown in Appendix Table 1.

shows the average management score of the control group enterprises.

[Insert Table 3 Here]

Panel A of Table 3 shows that the coefficients on the class+onsite group dummy and the class-only group dummy are positive and significant at the 1 percent level in the first and third follow-up surveys. The estimated coefficient on the onsite-only group dummy was only marginally significant in the first follow-up survey, but it is highly significant in the third follow-up surveys. As shown in column 5, these coefficients as a whole are highly significant. Similarly, Panels B to E indicate that the three training programs had generally significant impacts on the partial scores of the three treatment groups. Only exceptions are the impacts on the sales promotion and marketing scores of the onsite-only group. The salient finding from Table 3 is that the effects of the programs on management practices were felt soon and lasted at least for three years.[13]

An unexpected result is that the medium-run impact (i.e., the impact that appears in the third year) of the combined program on the management score is weaker than that of the classroom only program (see Panel A, column 3 of Table 3). This result is somewhat striking because the class+onsite group was invited to the on-site program as well as the classroom program, and because the impact of the on-site program is found positive and significant. We will discuss this issue after the next two subsections.

Throughout Panels A to E, the coefficient in column 2 has a lower significance level than the coefficients in columns 1 and 3. A reason why the significance level declines in the second data point may be that the control group mean of the management

13 Appendix D reviews how the adoption rate of each of the 27 practices varied over time.

score increased sharply from 12.8 to 17.5 probably due to knowledge spillovers. Indeed some sample entrepreneurs pointed out that the training participants and even non-participants initially tried to adopt as many practices as possible.

The same informants, however, said that they had abandoned some practices subsequently because they had found it tiresome to carry on those practices or to keep telling their workers to do so. Appendix Table 1 reports the percentage of the sample enterprises that were using each of the 27 practices included in the management score. As indicated by asterisk, between the second and third follow-up surveys, six practices were abandoned by more than 30 percent of the sample entrepreneurs. One was to spend money in advertisement, but the remaining five practices were *Kaizen* practices that would actually be implemented by workers rather than by entrepreneurs. Probably, the non-participants who imitated some practices would fail to learn from the participants how to encourage their workers to carry on the practices voluntarily. It would be more tiresome for the non-participants to maintain such practices. This is our conjecture about the reason why the most drastic decline in the management score occurred to the control group.[14]

C. Impacts on Business Performance

Figures 2 plots the cumulative distribution function (CDF) of the baseline value added by the treatment status. There are three outliers in the control group and one in the onsite-only group. This is why the baseline value added and sales revenues of these groups are greater than those of the other two groups, as shown in Table 1. Except for

14 Good practices, even after being used for a while, can disappear from workplace. Such disappearance does not seem uncommon. See, for example, Jackson and Schneider (2015) who find that garage mechanics stopped using checklists even though checklists work as a monitoring device and boost productivity.

these outliers in the upper tail, however, the CDFs of the four groups overlap considerably. Indeed, the Kolmogorov-Smirnov test does not reject the null hypothesis of equality in the baseline distributions between any two groups.

[Insert Figures 2 and 3 Here]

Panels A and B of Table 4 are similar in design to Table 3. As shown in the fourth column of these panels, the value added and sales revenues of the control group fluctuated considerably during the intervention and observation periods. Our respondents described year 2010 as a year of favorable market and ascribed the subsequent downturn to soaring material prices and a flood of low-priced imports into the Dar es Salaam market. Although all groups suffered, the control group was the most heavily affected, and the class+onsite group was the least affected. The latter's average value added and sales revenues did not sink below the baseline levels. Figure 3 plots the CDF of value added in 2013. The CDF of the class+onsite group is located to the right of the CDFs of the other three groups except for the upper tail, indicating that the combined program had a favorable effect on the value added of the treated. The Kolmogorov-Smirnov test rejects the null hypothesis of equal distributions between the class+onsite group and the control group in 2013 at the 5 percent significance level.

[Insert Table 4 Here]

In Table 4, Panels A and B have no significant estimates in the first two columns, which indicates that none of the three programs had significant impacts on business

performance in the first two years. In the next two years, however, the impacts of the combined program on both value added and sales revenue were significant. The magnitude of the impact of the combined program on value added in 2012 and 2013 is quite large compared with the corresponding average value added of the control group. These results, together with the results reported in Table 3, lend support to Hypothesis 1.

For the class-only and onsite-only groups, while the estimated impacts on value added and sales revenues were insignificant until 2012, those on value added turned to be marginally significant in 2013, which is consistent with Hypothesis 1. The significance levels are lower for these short programs than for the combined program, which lends support to Hypothesis 2, even though the magnitudes of the estimated impacts are not much different.

In Panel C, the outcome variable is employment size. Here we do not find any significant impacts of the training in the short or medium run. Panel D reports the results regarding investment in sewing machines and other capital goods, which is undertaken infrequently. We do not have baseline data on investment, but we collected recall data on investment in our second and third follow-up surveys.[15] The estimated impact of the combined program on investment is negligible in the earlier period, but it is positive and marginally significant for the later period even though the impacts of the short programs are insignificant. These results are consistent with Hypothesis 2.

Both classroom and on-site training programs covered *Kaizen* management, which emphasizes how to encourage or induce workers to reduce wasteful use of materials. For given sales revenue, reductions in wastes would increase value added. It

15 Because of the limited accuracy of recall data in general and the infrequency of investment, the data were transformed into a dummy variable that is 1 if any capital investment was undertaken during each of the two overlapping periods January 2011 - September 2012 and January 2012 - December 2013.

is expected that the training programs would have stronger impacts on value added than on sales revenues. Consistently, Panes A and B of Table 4 indicate that the estimated effect of each program as percentage to the control group mean is higher for value added than for sales revenues in 2012 and 2013, and the significance level for value added is not lower than that for sales revenue.[16]

D. Robustness Check

Because business performance data are generally noisy (de Mel, McKenzie, and Woodruff 2009), and because Figure 2 indicates that there are some outliers, we have attempted at winsorizing and trimming the top 5 percentile of the distribution of value added to check the robustness of the results obtained in Panel A of Table 4. As reported in Panels A and B of Tables 5, the estimated effect of the combined program is highly significant in 2013 (see column 4), even though its magnitude is substantially small compared with 13,163 dollars, the corresponding estimate in Table 4. Even in 2012, the estimated effect of the combined program is marginally significant in Panel A.

[Insert Table 5 Here]

As another robustness check, we included the record keeping score in the estimation of equation 1, following the lead of de Mel et al. (2014). The training participants may have become more meticulous in record keeping than the non-participants and may have come to provide the enumerators with more accurate

[16] These results are somewhat reminiscent of Anderson-Macdonald's (2014) finding on the differential impacts of marketing training and financial training on micro firms in South Africa.

information on revenues and costs. The effect of such a change may be controlled by inserting the record keeping score in the regression. As shown in Panel C, the results remain qualitatively the same as the corresponding results in Panel A of Table 4.

In Panel D of Table 5, distant recall data are excluded from the regression. The data on sales revenues and costs in 2008 were collected at the baseline survey in April 2010, and those in 2012 were collected at the third follow-up survey in January 2014. Suspecting that the quality of these distant recall data may be low, we ran regressions without using the 2008 data (i.e., y_{i0} is the value added in 2009, not the average of the 2008 and 2009 value added) and the 2012 data (which reduces the number of observations becomes smaller by 24 percent). The estimated impact of the combined program is significant in 2013 at the 5 percent level. Lastly, Panel E reports the results of taking the logarithm of the outcome variable. Although the estimated coefficients have lower significance levels, those of the combined program on value added in 2012 and 2013 are still marginally significant.

E. Discussion

The results that we have obtained lend support to Hypothesis 1; that is, a significant effect of a management training program on business performance is detected later than that on management practices. This hypothesis, however, is not specific about what the effect of training on management practices would get stronger or weaker over time. What we have found is that this effect became weaker in the medium run after becoming highly significant and large in magnitude in the short run (see Table 3). This finding is not surprising because the excitement and fervor of the participants may ebb away, because they may find it tiresome to implement practices or keep telling workers

to do so, or because they may find some practices unfit for their operation.

Nonetheless, the participants did not abandon all the practices that they had adopted after training. On the contrary, they kept some practices in use. Moreover, the positive and significant estimate of the medium-run effect of the combined program on value added (see Tables 4 and 5) suggests that these practices contributed to the improved performance. In interpreting these findings consistently, the key word seems to be assimilation. After the training programs, the participants would try many practices and then identify and focus on useful practices that were then assimilated. The favorable effect of training on business performance would come after sufficient assimilation.

Our management score is constructed based on binary data indicating whether a practice is implemented or not, whereas the management scores developed and used by Bloom and van Reenen (2010) and many others use data on the extent to which a practice is implemented properly. With binary data, an enterprises that has focused on and assimilated a small number of practices and implements them in a way that leads to high business performance can have a lower management score than another enterprise that implements a large number of practices in a way that leads to low performance. Probably because of this nature of our management score, the medium-run effect of the combined program on the management score appears weaker than that of the classroom only program in Panel A, column 3 of Table 3. We consider this unexpected result to be consistent with the result that the medium-run effect of the combined program on performance is stronger than that of the classroom only program in Tables 4 and 5.

IV. Conclusion

This paper has reported our follow-up observations for three years after a randomized controlled trial of management training. As intervention, we provided two short programs and a combined program for small enterprises. In line with the stylized finding of the existing studies, we have found that all the three programs improved management practices in the short run, but that none of the programs improved business performance in the short run. As a result of extending the follow-up observation period, we have also found that the favorable effect of training on management practices remained statistically significant even in the medium run, and that the combined training program had a significant effect on business performance in the medium run. The short programs had a marginally significant or insignificant effect on performance in the medium run. These findings suggest that the existing studies provided too little training or evaluated training impacts too soon to detect a favorable training impact on business performance.

Interestingly, when the participants in the combined program adopted many practices that they learned from the program, their business performance was not significantly better than that of the control group. Later they ceased to use some practices but outperformed the control group significantly. These findings suggest that assimilating new knowledge obtained from a training program is a necessary step toward improvement in performance.

There is a caveat regarding the comparison between our experiment and many existing ones. They differ not only in the length of follow-up observation period but also in two other respects. One is enterprise type. Many sample enterprises in many existing studies have no workers except for the self-employed owners and their family members. By contrast, many enterprises in our sample employ more than five paid workers and

are eager to expand their business sizes. Our experiment differs from the existing ones also in training content. Although ours included standard training content, it also covered the *Kaizen* approach to production management and quality control, which is expected to reduce the wasteful use of materials and other inputs and to improve the division of labor among workers or sections. Our experimental design does not allow us to assess the degree of contribution that each of these deviations from the existing experiments made to the new results. This issue, together with the issue of information spillovers, is deferred to future research.

Appendix A: Surveys

Our enumerators conducted personal interviews with entrepreneurs while filling out questionnaires. They basically used the same questionnaires in different survey rounds, although we deleted some questions that were found to be meaningless in the earlier survey. Among such questions, the most important one is about the willingness to pay for the training participation (see Table 2). At the interim survey in 2010, almost all the sample entrepreneurs, regardless of treatment status, answered affirmatively. At the first follow-up survey, all of them answered affirmatively. Thus, we did not ask questions about this issue subsequently. There was only one major addition to the questionnaire. At the interim survey, we added a new section on acquaintances and conversations with them about the training content (see Appendix C below), which was not included in the baseline survey questionnaire, and we expanded this section in the follow-up survey questionnaires by adding questions about visiting other enterprises and copying practices or layouts of other enterprises.

Our questions about management practices cover the 27 diagnostic criteria as listed in Appendix Table 1. For many criteria, visual inspection was used to judge whether the criteria were satisfied or not. However, for criteria such as “The entrepreneur has a clear sales or profit target in this year,” visual inspection is impossible, and direct questions such as “do you have a clear sales or profit target?” may lead to false answers. The questionnaire instead asks the entrepreneur to talk about his or her sales or profit target, and it asks the enumerator to judge whether the enterprise has set a clear target.

We measure business performance by value added and sales revenue. The questionnaire does not directly ask about the value added, even though de Mel et al.

(2009) recommend asking entrepreneurs directly about how much profit they earn rather than calculating profits. Our enumerators asked the entrepreneurs about the quantity and price of each of their products as well as those of each material used, the subcontracting costs, energy costs, and transportation and communication costs. Written records were also used whenever available. The enumerators then estimated the 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 the value added until the amount made sense to the entrepreneur. All the enumerators held bachelor's degrees and received survey training from us before each of the baseline and follow-up surveys so that they could handle such data collection procedures.

Appendix B: Baseline Correlations

The baseline management score is closely associated with the number of years of schooling and the number of years of business operation, as indicated by column 1 of Appendix Table 2. 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 the managers. The baseline value added and sales revenue are also correlated with the schooling years (see columns 2 and 4), but the correlation becomes weaker when we include the baseline management score as a right-hand side variable (see columns 3 and 5). The baseline management score is highly associated with both value added and sales revenue. In column 6, the dependent variable is the dummy variable that is equal to 1 if an enterprise was operating at the time of our third follow-up survey and zero otherwise. The regression result indicates that enterprise survival was not correlated with the treatment status.

Appendix C: Knowledge Spillovers

Appendix Table 3 reports the results of the 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 the training content. 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 in his or her social circle and is 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 a close association with the outcome variables when the influences of the other variables are controlled for. A major finding is that the most of the contact variables have positive coefficients, some of which are statistically significant. In column 1, the interaction of the number of contacts and the first and third follow-up survey dummies has a positive and significant coefficient. These results indicate that knowledge spilt over to acquaintances. Similarly, in columns 3 to 6, the interaction of the number of contacts and year dummies has a positive coefficient, and that with the 2010 and 2011 dummies has positive and significant one, suggesting that entrepreneurs with wider network have higher business performance. This may partially be due to the knowledge spillover but it is also due to the direct effect of having wider network in business, or sociability of entrepreneurs.

Another major finding is that in column 3 and 4, the impact of training on the

value added (as captured by the coefficient on the interaction of Class + Onsite and the 2012 and 2013 year dummy) is estimated to be positive and significant. This result seems to reinforce the results shown in Table 4 and the conclusion of this paper.

Appendix D: Management Score Components

Of the 27 practices constituting the management score, many were increasingly adopted by the training participants and non-participants. Appendix Table 1 shows the rate of adoption for each practice. Among the four sales promotion practices, three had a substantial increase in adoption by more than 20 percentage points from the baseline survey to the third follow-up survey. Two record keeping practices and four marketing practices also had such surges in adoption.

By contrast, only two out of the 15 *Kaizen* practices had such surges in adoption. Interestingly, however, the adoption rate increased by more than 20 percentage points by the second follow-up survey for eight *Kaizen* practices, but it fell sharply in the third follow-up survey. According to our interviews with respondents, this is because they initially adopted some practices but later found them not to be very useful. Their point may be illustrated by the changes in the adoption rates of practices 17 ("Every tool has a designated place"). The adoption rate of practice 17 increased from 35 percent in the baseline to 71 percent in the second follow-up survey, but it returned to 35 percent in the third follow-up survey. The sharp decline makes sense: Having a designated storage area for every tool would be meaningless if workers did not return the tools to their storage areas.

If, however, the storage areas for some tools were labeled clearly, workers are more likely to return those tools because it is clear to colleagues that the tools have not

yet been returned to their storage areas. The trainers taught the participants the trainers the usefulness of visualization by means of labeling and other devices when they taught the importance of returning tools to designated places. Many participants, however, did not fully understand the visualization part of the story and skipped it when they talked to their non-participant friends about the training program. Probably this is the reason why the adoption rate of practice 17 increased rapidly and dropped sharply.

Practice 18 (“Tools and their storages are labeled so that workers can easily find necessary tools”) was not as easy for the non-participants, who have never seen what kind of labels are used, to adopt. Even the training participants who missed the trainers’ explanation about visualization would not be able to adopt this practice. Thus, the adoption rate of this practice increased only slowly from 4 percent to 23 percent. But it remained at 19 percent at the third follow-up survey probably because many of those who once adopted this practice found it useful.

REFERENCES

Anderson-Macdonald, S. 2014. "The impact of marketing (versus finance) skills on firm performance: Evidence from a randomized controlled trial in South Africa," London Business School, *Job Market Paper*.

Barro, R. and 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., and 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., and Tungodden, B. 2014. "Human and financial capital for microenterprise development: Evidence from a field and lab experiment," *Management Science*, 61 (4), 707-722.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. 2013. "Does management matter? Evidence from India," *Quarterly Journal of Economics*, 128 (1), 1-51.

Bloom, N., Genakos, C., Sadun, R., and van Reenen, J. 2012. "Management practices across firms and countries," *Academy of Management Perspective*, 26 (1), 12-33.

Bloom, N. and 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., and Freeman, P. 2008. "Eliciting willingness to pay without bias: Evidence from a field experiment," *Economic Journal*, 118 (525), 114-137.

Bruhn, M., Karlan, D., and Schoar, A. 2017. "The impact of consulting services on small and medium enterprises," *Journal of Political Economy*, forthcoming.

De Grip, A. and Sauermann, J. 2012. "The effects of training on own and co-worker productivity: Evidence from a field experiment," *Economic Journal*, 122 (560), 376-399.

De Mel, S., McKenzie, D., and Woodruff, C. 2009. "Measuring microenterprise profits: Must we ask how the sausage is made?" *Journal of Development Economics*, 88 (1), 19-32.

De Mel, S., McKenzie, D., and Woodruff, C. 2014. "Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka," *Journal of Development Economics*,

106 (C), 199-210.

Drexler, A., Fischer, G., and 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.

Higuchi, Y., Nam, V.H., and Sonobe, T. 2015. "Sustained impacts of *Kaizen* training," *Journal of Economic Behavior and Organization*, 120, 189-206.

Hosono, A. 2017. "Industrial strategies: Toward a learning society for quality growth." In Noman, A., Stiglitz, J.E. (eds.) *Efficiency, Finance, and Varieties of Industrial Policy: Guiding Resources, Learning, and Technology for Sustained Growth*. New York: Columbia University Press.

Imai, M. 2012. *Gemba Kaizen: A Commonsense Approach to Continuous Improvement Strategy*, Second Edition. New York: McGraw-hill.

Jackson, C.K. and Schneider, H.S. 2015. "Checklists and worker behavior: A field experiment," *American Economic Journal: Applied Economics*, 7 (4), 136-168.

Karlan, D. and 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., and 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. and Puerto, S. 2017. "Growing markets through business training for female entrepreneurs: A market-level randomized experiment in Kenya" *World Bank Policy Research Working Paper* 7993.

McKenzie, D. and 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.

McKenzie, D. and Woodruff, C. 2017. "Business practices in small firms in developing countries," *Management Science*, forthcoming.

Morgan, J. and 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., and 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)
	Mean by treatment status				*p*-value for *t*-test		
	Class+ onsite	Class -only	Onsite -only	Control	(1) = (4)	(2) = (4)	(3) = (4)
Panel A: Control Variables							
Age (as of the baseline survey)	44.2	44.9	45.9	44.7	0.83	0.91	0.59
	(8.72)	(7.08)	(9.35)	(7.86)			
Male (yes = 1)	0.08	0.17	0.14	0.28	0.05	0.32	0.18
	(0.27)	(0.38)	(0.35)	(0.46)			
Chagga tribe (yes = 1)	0.15	0.25	0.34	0.28	0.25	0.80	0.60
	(0.37)	(0.44)	(0.48)	(0.46)			
Years of education	11.3	10.3	10.7	10.8	0.50	0.52	0.86
	(2.62)	(2.12)	(2.62)	(2.84)			
Prior business training	0.73	0.67	0.59	0.56	0.19	0.44	0.85
participation (yes = 1)	(0.45)	(0.48)	(0.50)	(0.50)			
Work experience as textile	0.15	0.25	0.24	0.19	0.74	0.58	0.61
employee (yes = 1)	(0.37)	(0.44)	(0.44)	(0.40)			
Years of operation (as of the	11.9	11.8	12.3	10.7	0.44	0.48	0.32
baseline survey)	(5.45)	(4.85)	(6.44)	(6.11)			
Number of employees (mean	5.5	4.9	5.7	4.3	0.29	0.62	0.30
of the 2008 and 2009 values)	(4.63)	(5.44)	(6.78)	(3.36)			
TANCRAFT member (yes = 1)	0.42	0.46	0.48	0.59	0.20	0.32	0.39
	(0.51)	(0.51)	(0.51)	(0.50)			
HOT member (yes = 1)	0.46	0.46	0.52	0.28	0.16	0.18	0.06
	(0.51)	(0.51)	(0.51)	(0.46)			
Panel B: Outcome Variables							
Management Score, 0 to 27	11.7	10.3	10.7	10.3	0.11	0.92	0.69
	(3.53)	(2.39)	(4.49)	(3.30)			
Value Added [GK$] (mean of	14473	13716	18665	27631	0.23	0.22	0.44
the 2008 and 2009 values)	(10964)	(12028)	(30876)	(54275)			
Sales Revenue [GK$] (mean of	23328	23130	28818	41549	0.23	0.25	0.43
the 2008 and 2009 values)	(15784)	(20723)	(45466)	(74714)			
Employment Size (mean of the	5.9	4.9	5.9	4.7	0.25	0.82	0.36
2008 and 2009 values)	(4.44)	(4.89)	(6.39)	(3.53)			
p-value for joint orthogonality					0.54	0.91	0.40
Number of enterprises in the group	26	24	29	32			
Number of participants in the Classroom training	22	23	0	0			
Number of participants in the On-site training	26	0	29	0			

Notes: Numbers in parentheses are standard deviations. Columns (5) to (6) show the *p*-values for the *t*-test of the null hypotheses that the two groups share the same mean values. Value added and sales revenues are expressed in the international dollar (i.e., the Geary-Khamis dollar) by using the PPP conversion factor provided by the World Bank's *World Development Indicators*. The *p*-value for joint orthogonality is obtained from the *F*-test of the null hypothesis that all coefficients are zero in the OLS regression of each treatment status dummy variable on the control and outcome variables.

TABLE 2—KNOWLEDGE SPILLOVERS

	(1) Class+ onsite	(2) Class-only	(3) Onsite-only	(4) Control	(5) Total
(a) Number of sample entrepreneurs with whom the respondent has talked about training content	12.2 (9.7)	11.6 (5.9)	7.9 (7.5)	4.0 (4.3)	8.6 (7.7)
(b) Number of training participants with whom the respondent has talked about training content	11.0 (8.6)	10.3 (5.1)	6.7 (6.5)	3.6 (3.9)	7.7 (6.8)
(c) Willingness to pay (yes = 1)					
(c.1) Baseline survey	0.65 (0.49)	0.75 (0.44)	0.66 (0.48)	0.69 (0.47)	0.68 (0.47)
(c.2) 1st follow-up survey	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)

Note: Numbers in parentheses are standard deviations.

TABLE 3—IMPACT ON MANAGEMENT SCORE (INTENTION TO TREAT -ANCOVA)

	(1) 1st follow-up	(2) 2nd follow-up	(3) 3rd follow-up	(4) *P*-value equality	(5) *P*-value all zero	(6) No.obs.
Panel A: Management Score [0-27]						324
Class+onsite	4.83	1.46	2.73	0.01	0.00	
	(0.86)	(0.81)	(0.82)			
Class-only	3.55	1.59	3.88	0.07	0.00	
	(0.90)	(0.71)	(0.89)			
Onsite-only	1.64	1.61	2.37	0.71	0.01	
	(0.94)	(0.73)	(0.87)			
Control group mean	12.8	17.5	13.1			
Panel B: Sales Promotion [0-4]						324
Class+Onsite	0.80	0.49	0.58	0.61	0.04	
	(0.30)	(0.31)	(0.27)			
Class-only	0.65	0.44	0.95	0.33	0.03	
	(0.32)	(0.33)	(0.31)			
Onsite-only	0.24	0.082	0.44	0.62	0.58	
	(0.27)	(0.29)	(0.32)			
Control group mean	1.5	2.6	1.5			
Panel C: Record Keeping [0-4]						324
Class+Onsite	0.81	0.24	0.96	0.01	0.00	
	(0.21)	(0.19)	(0.27)			
Class-only	0.65	0.27	1.00	0.01	0.00	
	(0.22)	(0.17)	(0.26)			
Onsite-only	-0.031	0.18	0.84	0.01	0.01	
	(0.32)	(0.20)	(0.26)			
Control group mean	3.1	3.6	2.9			
Panel D: Marketing [0-4]						324
Class+onsite	1.35	0.21	0.13	0.00	0.00	
	(0.31)	(0.16)	(0.27)			
Class-only	0.44	0.24	0.42	0.74	0.04	
	(0.34)	(0.16)	(0.20)			
Onsite-only	0.11	0.19	0.20	0.96	0.46	
	(0.33)	(0.17)	(0.20)			
Control group mean	2.2	3.5	3.1			
Panel E: Kaizen [0-15]						324
Class+onsite	2.20	0.87	1.39	0.21	0.00	
	(0.56)	(0.54)	(0.59)			
Class-only	1.87	0.82	1.63	0.42	0.00	
	(0.57)	(0.56)	(0.61)			
Onsite-only	1.38	1.24	1.04	0.91	0.01	
	(0.57)	(0.56)	(0.68)			
Control group mean	5.9	7.7	5.6			

Notes: Numbers in parentheses are standard errors clustered at the enterprise level. All regressions include the baseline value of the dependent variable, the enumerator dummies, and survey round dummies on the right-hand side, even though their estimated coefficients are not reported. Non-operating enterprises are excluded. There were no such enterprises in the first follow-up survey, but there were three in the second and six in the third follow-up surveys. *P*-value equality means the *p*-value for the test of the null hypothesis that the treatment effect remained constant over time. *P*-value all zero means the *p*-value for the test of the null hypothesis that the treatment effect is zero throughout the three follow-up surveys.

TABLE 4—IMPACT ON BUSINESS PERFORMANCE (INTENTION TO TREAT-ANCOVA)

	(1) 2010	(2) 2011	(3) 2012	(4) 2013	(5) *P*-value equality	(6) *P*-value all zero	(7) No. obs.
Panel A: Value Added [USD]							426
Class+onsite	-737.0	254.2	11014.0	13162.9	0.17	0.04	
	(7788.4)	(4963.5)	(5157.0)	(4868.3)			
Class-only	-5039.2	5118.1	6645.5	7241.0	0.72	0.57	
	(8212.6)	(5677.7)	(4775.6)	(4578.9)			
Onsite-only	4897.8	575.3	9116.6	10696.1	0.52	0.48	
	(16354.3)	(5194.4)	(6348.2)	(5983.0)			
Control mean	32367	25368	13281	13142			
Panel B: Sales Revenue [USD]							426
Class+onsite	-1930.7	2106.0	15866.9	19750.2	0.22	0.11	
	(9270.0)	(7043.6)	(7394.5)	(7442.8)			
Class-only	-3391.0	3395.5	8504.7	8798.0	0.83	0.76	
	(9784.4)	(7707.3)	(6588.2)	(6727.6)			
Onsite-only	3180.1	12242.4	11497.9	13259.9	0.94	0.42	
	(17530.3)	(8486.6)	(9081.2)	(8035.4)			
Control mean	47598	42219	26875	26947			
Panel C: Number of Employees							324
	1st follow-up	2nd follow-up		3rd follow-up			
Class+onsite	0.46	0.60		0.82	0.92	0.79	
	(0.92)	(0.96)		(0.81)			
Class-only	-0.18	0.14		0.35	0.85	0.95	
	(0.74)	(0.98)		(0.68)			
Onsite-only	2.47	3.52		1.68	0.24	0.23	
	(1.90)	(1.84)		(1.12)			
Control mean	4.6	5.3		4.2			
Panel D: Investment (yes = 1)							108 earlier
		Earlier		Later			105 later
Class+onsite		0.17		0.23	0.73	0.10	
		(0.13)		(0.12)			
Class-only		-0.041		0.19	0.13	0.24	
		(0.12)		(0.12)			
Onsite-only		-0.11		0.081	0.16	0.37	
		(0.10)		(0.10)			
Control mean		0.25		0.13			

Notes: Numbers in parentheses are standard errors clustered at the enterprise level. All regressions include the baseline value of the dependent variable, the enumerator dummies, and survey round dummies on the right-hand side, even though their estimated coefficients are not reported. Panel D is an exception, where the analysis is not of ANCOVA because the baseline data of the dependent variable are unavailable. Instead regression is run separately for two periods: the earlier period from January 2011 to September 2012 and the later period from January 2012 to December 2013. Non-operating enterprises are excluded. There were no non-operating enterprises in 2010, but there were three in 2011, nine in 2012, and six in 2013. *P*-value equality means the *p*-value for the test of the null hypothesis that the treatment effect remained constant over time. *P*-value all zero means the *p*-value for the test of the null hypothesis that the treatment effect is zero throughout the three follow-up surveys.

TABLE 5—IMPACT ON VALUE ADDED (ROBUSTNESS CHECK)

	(1) 2010	(2) 2011	(3) 2012	(4) 2013	(5) *P*-value equality	(6) *P*-value all zero	(7) No. obs.
Panel A: Winsorizing the top 5 percentile							426
Class+onsite	4571.4	-1263.0	5800.6	8135.7	0.23	0.08	
	(3844.3)	(4315.3)	(3029.3)	(2972.8)			
Class-only	2301.6	3583.5	1970.4	2127.1	0.99	0.88	
	(4085.3)	(5190.5)	(2447.5)	(2284.5)			
Onsite-only	-2244.3	2149.9	1043.0	3069.0	0.46	0.62	
	(3017.9)	(4243.4)	(2571.5)	(2877.2)			
Control mean	21227	22195	12749	12977			
Panel B: Trimming the top 5 percentile							402
Class+onsite	3639.4	-2544.5	3420.2	6274.3	0.25	0.16	
	(3704.0)	(3884.7)	(2191.1)	(2610.6)			
Class-only	-1131.8	69.8	2923.8	2585.1	0.62	0.58	
	(3432.8)	(4621.2)	(2156.3)	(2225.8)			
Onsite-only	-2828.6	2129.2	593.9	618.3	0.66	0.80	
	(3147.7)	(4307.4)	(1889.5)	(2268.9)			
Control mean	18439	19030	11687	11741			
Panel C: Controlling Record Keeping Score							426
Class+onsite	-688.7	295.8	11112.2	15866.3	0.18	0.09	
	(7760.4)	(4977.1)	(5224.1)	(6517.8)			
Class-only	-4987.5	5164.6	6692.1	10026.9	0.62	0.57	
	(8178.9)	(5686.6)	(4834.6)	(6313.3)			
Onsite-only	4930.0	601.7	9077.1	12990.2	0.49	0.52	
	(16350.5)	(5213.5)	(6361.9)	(7371.2)			
Control mean	32367	25368	13281	13142			
Panel D: Excluding Distant Recall Data							324
Class+onsite	926.0	2344.2		14964.2	0.10	0.02	
	(7222.0)	(5686.3)		(5777.8)			
Class-only	-3415.0	6768.5		8827.2	0.51	0.38	
	(7417.2)	(6106.2)		(5461.1)			
Onsite-only	6345.8	2229.8		12396.3	0.34	0.32	
	(15974.2)	(5639.6)		(6692.1)			
Control mean	32367	25368		13142			
Panel E: In Logarithm							426
Class+onsite	0.17	-0.035	0.35	0.37	0.42	0.38	
	(0.18)	(0.21)	(0.21)	(0.23)			
Class-only	0.055	0.036	0.14	0.19	0.94	0.89	
	(0.16)	(0.23)	(0.18)	(0.19)			
Onsite-only	-0.18	0.077	0.18	0.17	0.52	0.69	
	(0.20)	(0.25)	(0.21)	(0.23)			
Control mean	9.64	9.56	9.03	9.10			

Notes: Numbers in parentheses are standard errors clustered at the enterprise level. All regressions include the baseline value of the dependent variable, the enumerator dummies, and survey round dummies on the right-hand side, even though their estimated coefficients are not reported. Non-operating enterprises are excluded. The number of non-operating enterprises is 0 in 2010, three in 2011, nine in 2012, and six in 2013. In Panel D, the baseline value is not the average of the 2008 and 2009 values but the annual 2009 value. The *p*-value equality means the *p*-value for the test of the null hypothesis that the treatment effect remained constant over time. The *p*-value all zero means the *p*-value for the test of the null hypothesis that the treatment effect is zero throughout the four years.

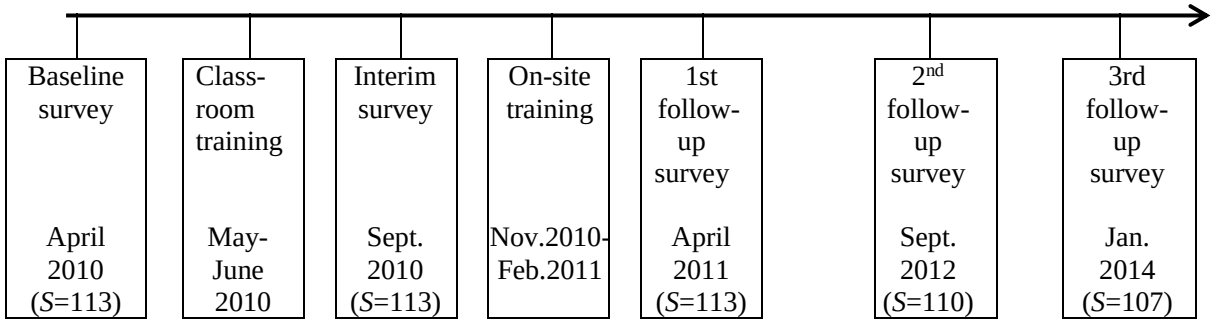

FIGURE 1. TIMELINE

Notes: S is the number of enterprises surveyed. Because two enterprises became model workshops and are excluded from the analysis, the size of the sample used in the analysis is $S - 2$.

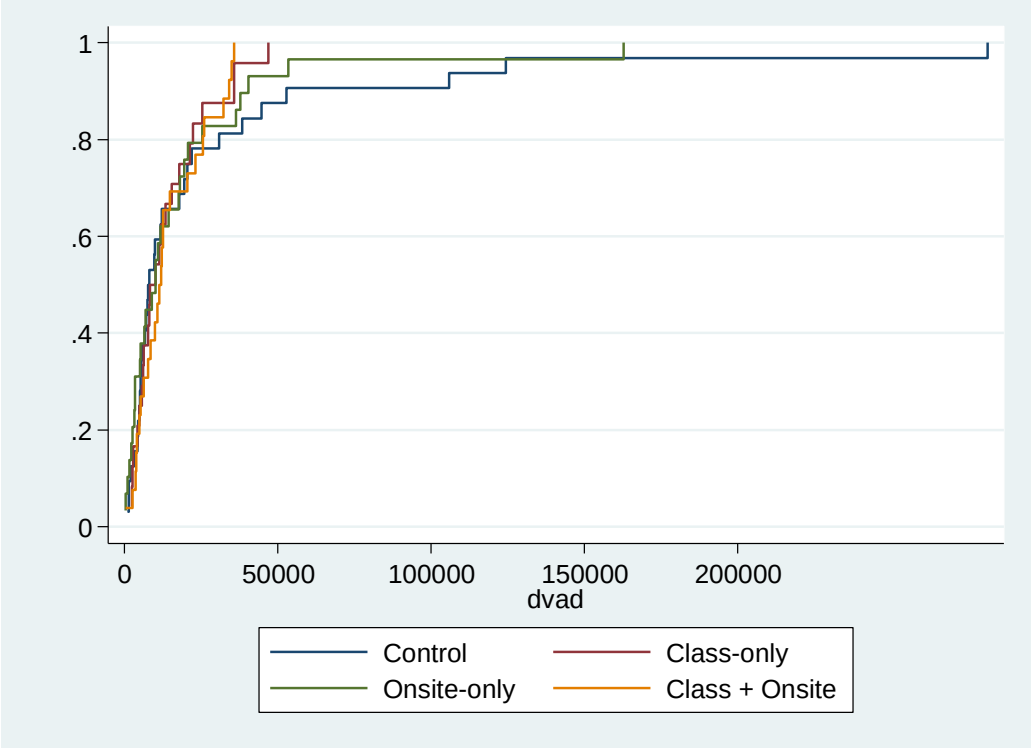

FIGURE 2. THE CDF OF THE BASELINE VALUE ADDED

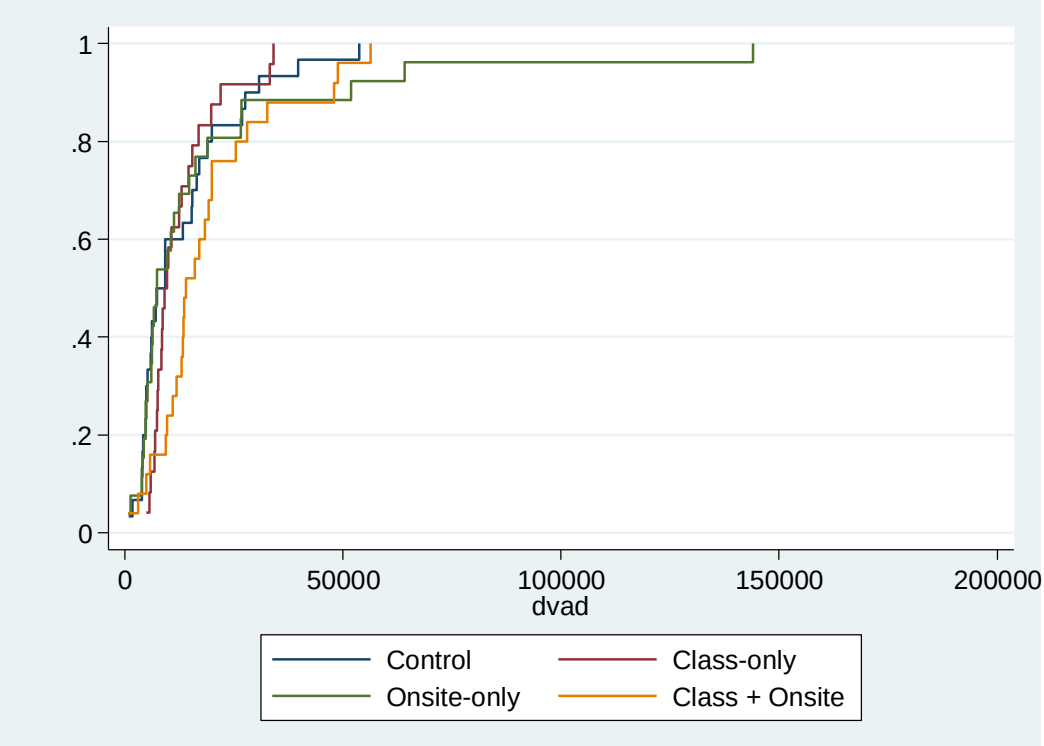

FIGURE 3. THE CDF OF THE 2013 VALUE ADDED

APPENDIX TABLE 1—MANAGEMENT SCORE COMPONENTS AND ADOPTION RATES (%)

Components	Survey Round			
	Base line	1	2	3
Sales promotion				
(1) The enterprise has had expenditure for advertisement in the last 3 months.	11	37	63	15*
(2) There is a signboard in front of the enterprise.	39	57	75	60
(3) The enterprise distributes complimentary cards or calendar.	25	43	80	57
(4) The enterprise issues invoices or receipts with its name or phone number.	37	59	77	62
Record keeping				
(5) Receipts and invoices are preserved.	49	81	96	92
(6) Business and household expenses are separated.	63	84	96	82
(7) Records of sales are kept.	85	92	97	93
(8) Records of material purchase are kept.	70	88	97	93
Marketing				
(9) The entrepreneur can clearly characterize his or her major customers.	43	67	93	85
(10) The entrepreneur can clearly describe its strength vis-a-vis its competitors.	24	62	88	93
(11) The entrepreneur has clear sales or profit target in this year.	45	73	96	67
(12) The entrepreneur has clear plan for its growth in the next five years.	28	62	90	92
Kaizen				
(13) A worker is assigned to inspect product quality before shipping.	11	5	3	2
(14) Records of quality defects are kept.	23	46	70	40*
(15) Records customers' complaints about the products sold are kept.	45	57	70	48
(16) The enterprise instructs the worker on the way to prevent defects.	9	2	4	8
(17) Every tool has a designated place.	35	53	71	35*
(18) Tools and their storages are labeled so that workers can easily find necessary tools.	4	11	23	19
(19) Every raw material has a designated place.	77	91	89	87
(20) Raw materials are stored separately from scrap or left-over materials.	76	93	94	83
(21) No left-over materials and cutting wastage are scattered on the floor.	14	62	61	56
(22) Scrap and wastage are removed and the floor is cleaned every day.	83	94	95	96
(23) Machinery is maintained at least once a week.	30	25	59	25*
(24) Regular meetings of all production workers are held.	28	48	64	53
(25) All production activities have their designated places.	30	38	52	22*
(26) The enterprise posts a flow chart describing the proper production process.	9	11	39	6*
(27) The entrepreneur is aware of the sequence and duration of each production process.	81	94	85	72

Notes: Numbers in the four columns on the right side are percentages of all sample enterprises.
* indicates practices that were abandoned by more than 30 percent of the sample entrepreneurs between the 2nd and 3rd follow-up survey.

APPENDIX TABLE 2—BASELINE CORRELATES OF MANAGEMENT SCORE, BUSINESS PERFORMANCE, AND ATTRITION (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline management score	Baseline value added		Baseline Sales revenue		Operating in the 3rd follow-up (1 = yes)
Age	-0.46	4444.1	6177.3	5394.2	7924.1	-0.019
	(0.50)	(4759.2)	(5133.6)	(5972.6)	(6628.0)	(0.023)
Age squared (/100)	0.42	-0.45	-0.61	-0.54	-0.77	0.023
	(0.57)	(0.51)	(0.54)	(0.63)	(0.70)	(0.023)
Male	-1.43	11634.9	17009.8	20211.1	28056.8	0.025
(yes = 1)	(0.95)	(16776.2)	(16410.2)	(22930.2)	(22024.9)	(0.042)
Chagga tribe	0.69	3800.6	1208.9	5668.4	1885.2	0.030
(yes = 1)	(0.77)	(5845.1)	(5552.2)	(8110.7)	(7408.7)	(0.046)
Years of education	0.31	2321.0	1151.8	3239.1	1532.4	0.013
	(0.11)	(1162.5)	(1049.6)	(1587.2)	(1500.8)	(0.0075)
Past business training	0.79	1779.9	-1198.4	6923.6	2576.2	0.030
experience (yes = 1)	(0.77)	(7838.6)	(7492.0)	(11304.0)	(10501.2)	(0.049)
Former textile employee	-1.04	600.8	4493.6	346.4	6028.7	-0.0036
(yes = 1)	(0.98)	(9351.8)	(8559.0)	(12790.6)	(11674.0)	(0.067)
Years of operation	0.18	789.0	113.0	1513.8	527.0	-0.0014
	(0.066)	(592.5)	(561.7)	(868.7)	(759.5)	(0.0053)
TANCRAFT member	-0.25	11052.8	11975.0	16813.3	18159.3	0.070
(yes = 1)	(1.11)	(7341.7)	(6789.0)	(10197.9)	(9300.3)	(0.11)
HOT member	0.38	922.4	-494.8	2358.8	290.2	0.023
(yes = 1)	(1.09)	(6044.0)	(5378.4)	(7927.7)	(6958.3)	(0.12)
Baseline mgmt score			3757.3		5484.4	
			(1368.1)		(1928.5)	
Class + onsite						0.028
						(0.064)
Class-only						0.078
						(0.052)
Onsite-only						-0.041
						(0.070)
No. observations	111	111	111	111	111	111
R^2	0.272	0.166	0.277	0.203	0.323	0.087
Adjusted R^2	0.174	0.054	0.172	0.097	0.224	-0.036

Notes: Numbers in parentheses are standard errors clustered at the enterprise level. All regressions, except for column 6, control for the enumerator fixed effects, even though their estimates are not reported. In column 6, the enumerator dummies were not included because non-operating enterprises were so few in number that the inclusion of the enumerator dummies led to insufficient variation for regression.

APPENDIX TABLE 3—SPILLOVER (INTENTION TO TREAT-ANCOVA)

	(1)	(2)	(3)	(4)	(5)	(6)
	Management score		Value added		Sales revenue	
	known	talked	known	talked	known	talked
Class + onsite			-5267.8	-8777.8	-8198.1	-12169.6
× 2010			(7648.0)	(8397.0)	(9409.1)	(9994.5)
Class + onsite	4.24	4.03	-3721.4	-3375.5	-4698.6	-4948.4
× 1st follow-up or 2011	(0.89)	(0.96)	(5035.5)	(5105.4)	(7499.1)	(7482.1)
Class + onsite	1.14	1.27	8490.8	9094.0	11515.8	11851.2
× 2nd follow-up or 2012	(0.76)	(0.76)	(5089.9)	(5458.9)	(7471.4)	(7683.6)
Class + onsite	2.14	1.95	11377.4	12105.3	16082.2	16647.1
× 3rd follow-up or 2013	(0.87)	(0.87)	(4747.8)	(5275.0)	(7514.9)	(7988.0)
Class-only			-10826.6	-12346.0	-11360.4	-12634.4
× 2010			(8767.3)	(9361.5)	(10819.7)	(11026.0)
Class-only	2.84	2.99	193.3	1783.8	-4940.3	-3060.3
× 1st follow-up or 2011	(0.92)	(0.89)	(6115.8)	(6189.3)	(8316.5)	(8283.0)
Class-only	1.15	1.33	3631.0	4939.9	3445.2	5097.7
× 2nd follow-up or 2012	(0.64)	(0.65)	(5134.7)	(5465.8)	(7281.0)	(7330.0)
Class-only	3.04	3.10	5054.8	6248.7	4470.5	6130.2
× 3rd follow-up or 2013	(0.94)	(0.94)	(4728.2)	(5157.4)	(7360.8)	(7514.0)
Onsite-only			3749.9	3404.7	336.2	252.8
× 2010			(17012.4)	(17312.0)	(18496.8)	(18749.0)
Onsite -only	1.19	1.33	-1424.6	-417.0	9235.7	10578.2
× 1st follow-up or 2011	(0.98)	(0.99)	(5033.4)	(5114.2)	(9061.4)	(9156.3)
Onsite -only	1.31	1.39	9046.9	9820.3	11092.8	12146.6
× 2nd follow-up or 2012	(0.69)	(0.69)	(6546.7)	(6698.3)	(9539.7)	(9573.8)
Onsite -only	1.93	2.00	9287.2	9971.5	10939.5	11969.4
× 3rd follow-up or 2013	(0.88)	(0.86)	(6224.0)	(6423.9)	(8569.7)	(8636.0)
No. contacts			398.4	1026.9	554.3	1301.0
× 2010			(239.7)	(510.9)	(259.0)	(538.5)
No. contacts	0.077	0.13	342.5	456.3	566.9	876.4
× 1st follow-up or 2011	(0.021)	(0.046)	(179.0)	(367.7)	(199.0)	(407.9)
No. contacts	0.010	-0.0086	200.1	214.3	346.2	467.9
× 2nd follow-up or 2012	(0.018)	(0.037)	(123.2)	(255.7)	(153.1)	(333.1)
No. contacts	0.053	0.10	141.9	108.9	295.5	357.3
× 3rd follow-up or 2013	(0.023)	(0.052)	(115.2)	(227.1)	(171.4)	(357.1)
No. observations	311	311	409	409	409	409

Notes: Numbers in parentheses are standard errors clustered at the enterprise level. All regressions control for the baseline values of each dependent variable, the enumerator fixed effects, and survey round dummies as explanatory variables even though their estimated coefficients are not reported.
Non-operating enterprises are excluded. There were no non-operating enterprise in 2010, but there were three in 2011, nine in 2012, and six in 2013.