Sustained Impacts of *Kaizen* Training☆

Yuki Higuchi[a], Vu Hoang Nam[b], Tetsushi Sonobe[c*]

a. *Nagoya City University, Nagoya, Japan*

b. *Foreign Trade University, Hanoi, Vietnam*

c. *National Graduate Institute for Policy Studies, Tokyo, Japan*

October 2015

Abstract

We conducted a randomized controlled trial of short-term management training for small manufacturers in two study sites in Vietnam and collected follow-up data repeatedly for two years to assess longer-term impacts than the existing studies of management training. Our training programs introduced participants to *Kaizen*, a common-sense approach to production management. In both sites, many participants started to recognize the importance of learning about management and improved their management skills. The impacts on management skills were statistically significant two years after the programs. Our results suggest that the training program increased participants' value added in one of the two study sites, likely because they learned how to eliminate wastes in production.

Keywords: Management training, Impact evaluation, Randomized controlled trial, Willingness to pay, Small and medium enterprises, Vietnam

JEL classification: L2, M1, O1

☆ We would like to thank Keijiro Otsuka, Alistair Munro, and Yasuyuki Todo as well as seminar participants at GRIPS and Nagoya City University for helpful comments. Very useful comments and suggestions have been received from two anonymous referees. This research was supported by the World Bank Japan PHRD Trust Fund No. TF096317 and by MEXT KAKENHI Grant Number 25101002.

* Corresponding author. Tel: +81 3 6439 6009. E-mail address: sonobete@grips.ac.jp

Highlights

- We assess the impacts of basic *Kaizen* training two years after the randomized trial.
- Positive and significant impacts on management skills were sustained for two years.
- Teaching concrete steps to eliminate wastes may be effective to increase business performance.
- The training increased willingness to pay for training participation.

1. Introduction

Managerial capital has increasingly been recognized by economists as a factor associated closely with enterprise productivity, growth, and longevity (e.g., Bloom and van Reenen, 2007, 2010; Bruhn et al., 2010; Syverson, 2011). The pioneering studies that conducted randomized controlled trials of management training have found that many small enterprises in developing countries are unaware of standard management practices common among their counterparts in developed countries, but that they adopt such practices after participating in a short-term training program (e.g., Berge et al., 2012; Bjorvatn and Tungodden, 2010; Bruhn and Zia, 2013; Drexler et al., 2014; Field et al., 2010; Karlan and Valdivia, 2011; Mano et al., 2012). Bloom, Eifert et al. (2013) find that a longer-term, on-site coaching program improved not only management practices but also business performance of medium-sized textile plants in India.

In their survey of the earlier studies of short-term management training for small enterprises in developing countries, McKenzie and Woodruff (2014) argue that this line of research has yet to provide useful information for policy makers. Indeed, the existing studies do not provide strong evidence that management training improves business performance in terms of such accounting-based indicators as value added. While this could be attributed to noisy data, small sample sizes, and inadequately designed training programs,[1] it is possible that the standard management practices in developed countries are not useful for small enterprises in developing countries and, for that reason, have seldom been adopted or known by them. Moreover, these studies,

1 For example, Bruhn et al. (2010) find that the estimated effects of their training intervention on productivity were "economically large but are only significant at the ten percent level." The authors ascribe these results to their noisy data and small sample size. Except for Berge et al. (2012) and Drexler et al. (2014), few studies examine the relative effectiveness of different training program designs, which may depend on trainees' types or needs.

except for Berge et al. (2014) and Karlan and Valdivia (2011), evaluate training impacts only in several months or a year after training interventions and cannot determine how long the training impacts last.

This paper attempts to address some of these issues by providing two training programs in two study sites and by conducting a baseline survey and three follow-up surveys for three years. We measured management practices, business performance, and willingness to pay for training participation, even though the training was provided for free. After a classroom training program, we conducted the first follow-up survey, which was followed by the provision of an on-site training program in which instructors visited trainees' enterprises two or three times. The second and third follow-up surveys were conducted soon after the on-site training and two years later, respectively, which allows us to estimate relatively long-term training impacts. The experiment was conducted with small and medium enterprises (SMEs) in two industrial clusters near Hanoi, Vietnam, one producing rolled steel construction materials and the other producing knitwear garments. SMEs in each cluster are relatively homogeneous as they share equal access to the same technology, the same product and intermediate input markets, and the same labor market. The short-term impacts of the same training program in one of the two clusters are evaluated by Suzuki et al. (2014), but their study was completed before the third follow-up survey was conducted. The present study is the first to estimate longer-term impacts by using third follow-up survey and also the first to compare the training impacts between the two clusters.

Another feature of our training intervention is that it provided introductory courses to the *Kaizen* approach to production management, which was developed in Japan based on the US-born industrial engineering and quality management ideas but now is widely accepted across the world as a standard approach (Imai, 2012). This approach encourages workers to spot inefficiency problems, such as uneven workflow, waste motion, inefficient workplace layout, and other inefficient practices and arrangements and to find solutions to the problems. While *Kaizen* is common among large firms in Vietnam, it was not known by SMEs in our study sites. Before the training programs, only a few SMEs were willing to pay a small amount of money for participation in our training programs, and the actual take-up rate for the classroom training was low.

Three major findings stand out. Firstly, our training programs, despite short sessions, had favorable and persistent effects on the production management of the trained enterprises which lasted for at least two years in both sites. This finding allays the concern about the sustainability of improved management practices. Secondly, while the impacts of the training on business performance are not always significant, our study suggests which part of the training program was useful to improve business performance. The useful part was the concrete steps to reduce dead stock taught in the training sessions even though other parts of the training might also contribute to performance improvement. Thirdly, willingness to pay, which was very low initially, increased significantly among training participants, suggesting that not a few participants found the training programs useful. As one might wonder how reliable our measures of management skills and willingness to pay are, the paper presents suggestive evidence that these measures reflect to some extent the true management skills and demand for training. Overall, the findings suggest that small enterprises seldom receive management training simply because they do not know the value of learning about management, but that training participation improves their perception as well as management skills. The training participation also has the potential to improve business performance if the level and intensity of training are appropriate.

The remainder of this paper is organized as follows. Section 2 describes the experimental design and Section 3 checks the balance and discusses the attrition. Section 4 presents the results of impact evaluation. Section 5 contains the summary of the findings and implications for future studies.

2. Experimental Design

2.1. Study Sites and Sample Enterprises

There are two thousand village industrial clusters throughout Vietnam which have spontaneously developed and produced traditional craft items, and some of them now produce modernized products as well as intermediate inputs for industries, according to JICA's (2004) survey of clusters in this country. The clusters have contributed to the rapid economic growth since 1986 when the economy was liberalized by *Doi Moi* (Renovation) policy (Oostendorp et al., 2009). Nam et al. (2009, 2010) selected two of these clusters for their enterprise surveys in 2007 as clusters that have successfully started the production of modern items. We chose these two clusters as our experiment sites partly because of the existing rapport, and partly because they are representative clusters of modern products in semi-urbanized areas in Vietnam in terms of the number of firms, the employment size per firm, and some other aspects.[2]

---

2 According to our interview with the lead consultant of Japan International Cooperation Agency's (JICA) project called "Artisan Craft Development Plan for Rural Industrialization in Social Republic of Vietnam," these two industrial clusters have a little greater number of customers than average because they are located relatively near the capital city, but they were neither particularly large nor advanced. The project is summarized in JICA (2004). For the details of the development process of the two industrial clusters, see Nam et al. (2009, 2010). Note that

The two clusters under study are located in suburbs of Hanoi about 15km from the city center but in different directions: one cluster in Bac Ninh province has produced steel products and the other in Ha Tay province has produced knitwear and garment products.[3] The steel cluster has a history of several centuries as a blacksmith village, even though the production of rolled steel products, the current main products, began only a few decades ago. The knitwear cluster was formed in the early 1990s by workers who had produced towels at a cooperative until it was closed in 1991. In the clusters, employees and subcontractors of enterprises are from the same or neighboring villages, but none in the neighboring villages produce final goods. Thus, the geographical area of each cluster coincides with the village border.

In the steel cluster, Nam et al. (2009) surveyed 204 enterprises randomly selected from the 372 enterprises that were in a list provided by the commune government office in 2007. In the knitwear cluster, Nam et al. (2010) surveyed a total of 138 enterprises in operation in 2007, even though the collected data were lost due to an accident in late 2008. The total number of knitwear enterprises was 161 according to the list complied in 2010 by the commune office. Almost all of our sample owners have only one enterprise. They have parents, siblings, parents-in-law, siblings-in-law, and other relatives engaged in the same trade. Children of business owners tend to start their own enterprises upon marriage by learning business skills and receiving financial assistance from their parents and in-laws. Some workers are employed on a time-rate basis, but many workers prefer to be piece-rate workers so that they can choose to

Vietnam has large-scale producers of knitwear and steel, including state-owned enterprises and foreign ventures, as well and that they are not located in these village clusters.

3 The knitwear cluster used to be located in Ha Tay province, which has been integrated into Hanoi since 2008, when Hanoi area was expanded.

engage in farming or work at an enterprise or two flexibly if needed or desired. Thus, it seems that information spillovers can be rampant because of the geographical density of enterprises, the kinship of enterprise owners, and the mobility of workers among the enterprises.

Just before our management training programs started in 2010, baseline surveys were conducted in the two clusters. We found that 155 of the 204 enterprises in the sample used by Nam et al. (2009) were still in operation in the steel cluster. Our baseline sample consists of these 155 steel enterprises and all the 161 knitwear enterprises registered with the commune office.

2.2. Experimental Intervention

In most existing studies, management training interventions highlight basic skills in accounting, marketing, and business strategy (McKenzie and Woodruff, 2014). Production management including quality control is seldom in the scope of these training programs. By contrast, our intervention features the *Kaizen* approach to production management.[4] According to our personal interviews with several business consultants, this approach works well in helping particularly those enterprise owners who are clueless about how to induce or encourage workers to keep workplace neat and tidy, to keep track of inventory and work in progress, to perform machine maintenance regularly, and to undertake other activities that would prevent injury, improve work

4 Van Biesebroeck (2003) examines the changes in the productivity of the U.S. automobile industry from the 1960s to the 1990s and finds that the introduction of the lean manufacturing and *Kaizen* considerably increased the productivity during the 1980s and 1990s. Garment enterprises in Bangladesh grew fast partly because of their engagement in *Kaizen* (Mottaleb and Sonobe, 2011). There have been few attempts to evaluate the impacts of *Kaizen* training in developing countries by means of randomized controlled trials. An exception is Sonobe and Otsuka (2014).

environment, decrease waste and increase efficiency. Such enterprise owners seem to abound in a number of industrial clusters in developing countries including our study sites. The *Kaizen* approach does not require sophisticated knowledge or capital investment to improve production management.

While *Kaizen* is said to be applicable to a variety of manufacturing sectors (Imai, 2012), the impacts of the *Kaizen* training may vary between the two clusters under study because steel production is more capital-intensive than garment production. For example, while *Kaizen* instructors encourage enterprises to create a smooth, safe, and efficient workflow, steel enterprises cannot, unlike knitwear enterprises, change workshop layouts on a trial and error basis to find out such a workflow because their machine and equipment are bulky and heavy. On the other hand, labor-intensive knitwear enterprises can mobilize workers and easily make small changes to find out better workflow to reduce various wastes in production. We suspect that the same management training have different impacts on experiment subjects in different sectors with different factor intensity. To our knowledge, this is an open question in the literature on the randomized controlled trials of management training.

We made a contract with a business consulting firm in Japan to dispatch a *Kaizen* expert to our study sites. We also hired a local consultant, who was qualified as a master trainer of the International Labor Organization’s (ILO) Start/Improve Your Business (SIYB) training, and her co-trainer. The *Kaizen* expert taught the local consultants in English, and the latter taught in the local language the training participants, who were enterprise owners or their adult children.

Bloom, Eifert et al. (2013) find that an extensive training program featuring lean manufacturing, an American version of *Kaizen*, was effective in improving management practices and productivity at medium-sized textile plants in India. It remains an open question whether less expensive, shorter-term training programs can have favorable and sustained impacts on small-sized enterprises. In our experiment, training intervention had two components: classroom training for 2.5 hours a day, five days a week over a three-week span, and on-site training for two to three full days. The timeline of each training program is shown in Table 1.[5] In the classroom training program, the local consultants taught standard contents of business development service, such as entrepreneurship, business strategy, marketing, and bookkeeping in addition to *Kaizen*.[6]

From among the classroom training participants, the team of instructors selected two enterprises in each cluster to make them model enterprises, which provided trainer training grounds and served as showcases of *Kaizen* practices. At the selected four enterprises, the instructor team convinced the owners to change the layout of their workshops. While doing so, the *Kaizen* expert taught the local consultants how to coach owners and workers on site. Subsequently, a one-day seminar was held in each cluster to let the model enterprise owners give presentations about their enterprises' physical changes and the responses from their workers as well as their own opinions. In the on-site training program, the local consultants visited participants' enterprises two or three times to demonstrate how to encourage workers to improve their work environment, productivity, and product quality. We provided two types of training

5 As the business owners in the steel clusters had not been willing to participate, we provided the same classroom training program twice in June-July 2010 and September 2010 in this cluster. 31 enterprises participated in the first classroom training session and 13 enterprises in the second one. In the empirical analyses, we treat enterprises as classroom training participants without distinguishing the first and second sessions. In the knitwear village, we provided the classroom training only once, in July - August 2010.

6 The *Kaizen* part of the training program included 5S, a workplace organization method consisting five activities: sorting, setting in order, shining, standardizing, and sustaining (or self-discipline).

programs to our sample enterprises so that we can compare the relative effectiveness.

2.3. Randomization

Stratified by the cluster, we randomly assigned the total of 316 enterprises in our baseline samples into three treatment groups and a control group. The first treatment group was invited to both classroom and on-site training programs and labeled as Group TT, while the second and third were invited only to either the classroom or the on-site program and labeled Group TC and Group CT, respectively. Group CC, the control group, was invited to neither of the programs. The model enterprises mentioned above, however, were not selected randomly. This is because, to serve as a model, an enterprise had to be willing to accept other training participants' visits and hence had to have enough space. Because they were treated differently from any other sample enterprise, we exclude them from our statistical analyses below, which leaves 312 enterprises in the sample (153 in the steel cluster and 159 in the knitwear cluster). While the assignment of enterprises into treatment and control groups was random originally, it is no longer completely random after the elimination of the model enterprises. The resulting sample size of each group is shown in the bottom of Table 2.[7]

While 108 enterprises in the steel cluster were invited to the classroom training program, only 41 enterprises actually participated. In the knitwear cluster, 89 enterprises were invited, and only 52 enterprises actually participated. We issued a

---

7 We provided the on-site training to a smaller number of enterprises due to the budgetary constraints and the availability of the instructors. Given the large variance in business performance at baseline, the number of samples needed to detect 10 percent change in business performance with 90 percent statistical power is over one thousand, which is far beyond our training and survey budget. Thus, we conducted a number of follow-up surveys as well as frequent visits to the study sites to collect qualitative information.

certificate to the enterprises that participated for at least ten days of the classroom training out of the total 15 days. We regard only the certificate recipients as classroom training participants.[8] The take-up rate was 38 percent and 58 percent in the steel and knitwear clusters, respectively. By contrast, the take-up rate of the on-site training was 100 percent in both clusters because no enterprise refused to accept the consultants' visits. There were no uninvited participants in any training program.

3. Balance Check and Attrition

3.1. Balance Check

Panel A of Table 2 presents the owners' characteristics by treatment status and by cluster, highlighting the differences between each treatment group and the control group. Columns 1 to 3 show the coefficients on the group dummies and the corresponding *t*-statistics obtained from the regression of each characteristic on the three group dummies (with Group CC being a reference group) and no other explanatory variables, in the steel cluster sample. Column 4 shows the means and medians of Group CC in the steel cluster. Column 5 reports the *p*-values of the *F*-tests of the null hypothesis that all the coefficients on the group dummies are zero. Similarly, columns 6 through 10 report the knitwear counterparts of these statistics. Another point to note on Panel A is that only 24 sample entrepreneurs (four in the steel cluster and 20 in the knitwear cluster) had received business training prior to our intervention. Of these entrepreneurs, ten received training in management in general, eight in accounting, three in human resource management, and three in information technology.

8 In Table 5, we present the results, where we treat samples participated in at least one day of the classroom training but did not receive the certificate as the classroom training participants. Such samples include six knitwear enterprises participated for less than six days, and four steel and ten knitwear enterprises participated between six to nine days.

Panel B presents data on the baseline level of management skills. To rate basic skills in production management, we constructed an indicator, which we call the *Kaizen* score, based on 11 diagnostic criteria listed in Appendix Table A1. During the survey, our enumerators visited each sample enterprise and judged whether the enterprise met each criterion based on either the enumerators' visual inspection or the owner's way of responding to their questions. The *Kaizen* score of an enterprise is the number of the diagnostic criteria that the enterprise was found to meet, and, hence, the lowest possible value is zero and the highest is 11. The score should be high if *Kaizen* is well established. Because *Kaizen* is a common-sense approach, some enterprises may have adopted some *Kaizen* practices and get somewhat relatively high scores without knowing that those practices are part of *Kaizen*.[9]

In order to measure the value that business owners attached to management training, our enumerators, after outlining our training programs, asked them a hypothetical question, "Would you pay three million Vietnamese Dong (VND) (about 150 USD as of 2010) to participate in the training program?" which was followed, if the answer was affirmative, by an additional question, "How sure are you about the answer? Are you definitely sure or probably sure?" According to the results of experiments by Blumenschein et al. (2008), the number of the "definitely sure" answers to a hypothetical question is close to the number of affirmative answers to the corresponding question associated with actual payment of the amount of money. Thus, from the answers to the above set of questions, we constructed a dummy variable that is 1 if the answer is "definitely sure" and 0 otherwise. We refer to this variable as WTP.

9 For example, Group CC in the steel cluster had an average score of 6.1, as shown in column (4) in Table 2, which is higher than the average score of Group TT in the knitwear cluster after the classroom training, as shown in Table 3.

Panel B of Table 2 presents the baseline values of annual value added, sales revenue and employment size (i.e., the number of workers) as well. The data on these baseline values are recall data collected in 2010. For the knitwear enterprises, the baseline values are the averages of real annual values in 2008 and 2009.[10] The average is taken to reduce noise in the data, following the lead of McKenzie (2012). For the steel enterprises, the baseline values are real annual values only in 2008.[11] As an indicator of business performance, we are more interested in value added than sales revenue, as the *Kaizen* approach emphasizes cost reduction through the elimination of wasted materials and time.[12]

In Panels A and B of Table 2, the null hypothesis of the *F*-test is rejected only for the baseline WTP (see Panel B, column 10) at the ten percent significance level but not for the other variables. The baseline WTP in the knitwear cluster varies across treatment groups possibly because of the small sample size. Aside from this, the randomization was generally successful.

3.2. Attrition

The first follow-up survey, which was conducted soon after the classroom training program, collected data on management practices and willingness to pay for training

10 The producer price index reported by General Statistics Office of Vietnam (GSO) was used as a deflator to obtain real values for value added and sales revenue. More precisely, the price index for metal products was used for the steel cluster and that for garments was used for the knitwear village. The deflator for the steel village is 1 for 2008, 1.014 for 2010 and 1.200 for 2012, while that for the knitwear village is 1 for 2008, 1.068 for 2009, 1.131 for 2010 and 1.421 for 2012. 1 million VND was approximately equivalent to 61 USD in 2008.

11 We failed to collect 2009 data partly because we did not know McKenzie's (2012) paper when we conducted the baseline survey in the steel cluster.

12 We did not directly ask the respondents how much the value added was. Instead we asked sales revenue, material costs, subcontracting costs, energy costs, and transportation costs. In the analyses below, value added is used as a main outcome variable capturing business performance, but those results obtained when sales revenue is used in place of value added are also reported.

participation but not sales revenue and production cost. The second follow-up survey was conducted soon after the completion of the on-site training, to collect data on management practices, willingness to pay, sales revenues and production costs. There was no incidence of attrition at the second follow-up survey. At the third follow-up survey, which was conducted in January 2013, however, we found that 25 steel enterprises and 12 knitwear enterprises had not been in operation throughout 2012. In addition, a knitwear enterprise in Group CC refused to cooperate with our data collection interview in the third follow-up survey.

A question arises as to whether the incidence of attrition is correlated with treatment status. Appendix 1 addresses this question. As Appendix Table A2 shows, whether or not sample enterprises could be traced in the third follow-up survey depends on treatment status, particularly in the knitwear cluster. Thus, we have to pay attention to possible attrition biases, and this is why we conduct bounds analysis in Appendix 1.

4. Empirical results

4.1. Descriptive Analyses

Table 3 shows the data on the *Kaizen* score, WTP, and value added measured after the training programs in the same fashion as Table 2: columns 1 to 5 are about the steel cluster while columns 6 to 10 are about the knitwear cluster; columns 1 to 3 and 6 to 8 show the regression coefficients on the treatment status dummies and their *t*-statistics; columns 4 and 9 show the mean and median in Group CC; and columns 5 and 10 show the *p*-value of the overall *F*-test. Soon after the classroom training, the *Kaizen* score of Group TT was significantly higher than that of Group CC in both clusters. Similarly, in both clusters, soon after the on-site training, the score of Group CT was higher than that of Group CC, and the score of Group TT became even higher. Moreover, while the coefficients were not significant for Group TC soon after the training, they became significant two years after the training in the knitwear cluster. By contrast, the score of Group CC remained almost constant. This constancy is an unexpected finding because it suggests that knowledge spillovers from the training participants to non-participants were not rampant but rather limited despite the dense human exchange in the clusters as described in Section 2.[13] As the *p*-values in columns 5 and 10 indicate, the deviations between the treatment groups and the control group in the *Kaizen* score two years after the training are highly significant in both clusters.

If knowledge spillovers are limited, it may be explained partly by the low level of WTP. This is also an unexpected finding. Very few enterprises were willing to pay 150 USD for the training. In the steel cluster, the mean value of this dummy variable is 0.03 before the training (see column 4 in Table 2) and 0.06 after the training programs (see column 4 in Table 3), which means only one and two sample enterprise owners, respectively, in the control group were willing to pay such a small amount. It may well be that enterprise owners in the control group remained uninterested in *Kaizen* even though they could have imitated those new management practices adopted by the training participants. Their indifference to *Kaizen* may be reasonable because the real value added of the enterprises in the treatment groups did not increase much relative to that of the control group enterprises. Although the difference between Groups TT and CC became significant at the 10 percent level in the steel cluster in 2010, the difference

13 Bloom, Eifert et al. (2013) document that some of the practices were transferred to other plants owned by treated enterprises of their study. This is unlikely to be the case in our study sites because few business owners have multiple enterprises or workshops. But many business owners are relatives, and workers can move from one enterprise to another. Thus, it is surprising to find that spillovers were limited.

became insignificant again in 2012. In the knitwear cluster, the real value added of the control group did not differ from that of the treatment groups significantly, even though the significance level increased in 2012 for the difference between Groups TT and CC and for the difference between CT and CC.

WTP, however, increased among those who were treated. In both clusters, the deviations between the treatment groups and the control group after the on-site training are statistically significant at any conventional level. Note that the coefficients on the treatment status dummies are greater in the knitwear cluster than in the steel cluster for both the *Kaizen* score and WTP. To the extent that these coefficients reflect the training impacts on WTP, these observations are consistent with the hypothesis that the training impact is stronger in the knitwear cluster than in the capital-intensive steel cluster, where the adoption of *Kaizen* practices is expected to be more difficult because of bulky and heavy equipment, materials, and products.

We admit that both WTP and value added can be prone to social desirability response bias, to use the terminology of Arnold and Feldman (1981). In other words, our respondents might exaggerate their WTP, a measure based on hypothetical questions, and their business performance, in order to please our enumerators. It is also possible that the enterprise owners in the control group, who were invited to neither program, pretended that they were indifferent to the training and were performing well. We suspect that Groups TT and CT in the steel cluster exaggerated their sales revenues to please or impress our enumerators in 2010, when the real estate market was booming and boosting the demand for construction materials including rolled steel bars and wires. In 2012, after the boom, the differences in real value added among the four treatment status groups disappeared in the steel sample. In the knitwear sample, however, the coefficients on the three treatment status dummies are greater in 2012, and the *p*-value of the overall *F*-test is almost 10 percent.[14] The difference between Groups CT and CC became particularly large in 2012, even though it falls short of the 10 percent significance level (see column 8 of Table 3).

At the bottom of Table 3, we present data on the percentage of unsold output in 2012, our measure of dead stock, in the knitwear cluster. Before the knitwear sales season, producers accumulate the stock of finished product. During the production period, they continue to produce while receiving orders from wholesalers. The problem is that they often continue to produce without knowing inventory levels by product, size, and color and end up with a large stock of unsold products, a complete waste of time and money. This overproduction problem arises partly because of their lack of proper product inventory control and partly because they tend to have a large inventory of work in progress at different stages of production process, which makes it difficult to predict future output. For steel product producers, unsold products do not cause a serious trouble because their products are much more durable. We collected the data on unsold products only in the third follow-up survey. The data indicate that while Group CC had on average 7.2 percent of their output unsold, the treatment groups sold almost all output by the end of the calendar year. In the steel cluster, although many enterprises, especially those in Group CC, would suffer from various kinds of waste other than unsold products, it was practically impossible for us to capture their waste problems numerically.

According to our open-ended interviews with workers at our sample enterprises,

14 The significance level is lower if value added data are replaced by sales data. This is consistent with the fact that *Kaizen* training emphasizes cost reduction.

workers in the steel cluster benefitted from the training because of reduced risk of injury. In the knitwear cluster, piece-rate workers, who account for the vast majority of workers, benefitted from *Kaizen* practices even though their piece rates remained the same. This was because they came to produce more and earn more. Even when they could not produce more due to the limited supply of materials, they could finish their work earlier and use the saved time for leisure or other income-generating activities.

4.2. Reliability of Outcome Measures

We admit that there are reasons to be skeptical about our measures of outcome variables. This sub-section is devoted to checking their reliability. In Table 4, columns 1 to 4 present the results of the regressions linking the baseline values of *Kaizen* score, WTP, and value added with the variables representing the owner's characteristics. Column 1 shows that the baseline *Kaizen* score is significantly associated with schooling in both clusters and with prior business training experience in the knitwear cluster. These results are consistent with Bloom and van Reenen (2010), who find that their measure of management practices is closely associated with the human capital of managers in data from a large number of medium-sized firms in developed and emerging economies. Similarly, in the steel cluster, WTP and the log of value added as of the baseline survey are associated with schooling.[15] In the knitwear, the baseline log of value added is associated with prior training experience. While admitting that the *Kaizen* score and value added are endogenous, we regress baseline value added on baseline *Kaizen* score in column 4 and find that they are closely

15 As not a few sample enterprises had negative value added, we use a transformation, $log\{y + (y^2 + 1)^{0.5}\}$, following the lead of Burbidge et al. (1988), and refer to it as the log of *y* for convenience.

associated with each other in the steel cluster, even though the corresponding coefficient in the knitwear cluster is insignificant. These results suggest that the *Kaizen* score reflects management skills to some extent.

As mentioned earlier, many enterprises in Groups TT and TC did not participate in the classroom training program. The decision on whether to participate was left up to these invited enterprises. The *ex ante* value of WTP should be a good predictor of this decision if WTP reflects the true demand for training. Thus, we estimated a probit model explaining participation in the classroom training at least ten days out of 15 days. The estimation results are presented in columns 5 and 6. It is clear from the table that *ex ante* WTP is by far the best predictor of the decision to participate in the classroom training in both clusters.[16] Overall, these results suggest that WTP reflects the true demand for training to some extent, and that the *Kaizen* score and value added are reasonable measures of management skills and business performance.

4.3. Econometric Specification

We first estimate the sustained, or relatively long-run, impacts of the training, without distinguishing the classroom and on-site training programs to increase the statistical power for detecting training impacts. Let y_i be an outcome variable, which can be the *Kaizen* score, WTP or the log of value added, of enterprise *i* at the latest data point, that is, the third follow-up survey for the *Kaizen* score and value added, and the second follow-up survey for WTP. Our basic specification may be written:

16 In addition, if our WTP variable is reasonable as a measure of the true demand for training, those training participants with increased WTP are expected to improve their management practices and business performance. Indeed we have a table indicating such relationship, which is available upon request to the corresponding author.

$$y_i = \alpha P_i + \boldsymbol{X_i\gamma} + \varepsilon_i , \quad (1)$$

where P_i is a dummy variable indicating whether enterprise i participated in at least one of the two training programs, $\boldsymbol{X_i}$ is a vector of variables capturing the entrepreneurs' essentially time-invariant characteristics,[17] and ε_i is an error term. Under this specification, α is our measure of the sustained impact of the training.

To estimate the impact of each training program at each data point, equation (1) may be modified as follows:

$$y_{it} = \sum_{s=1}^{3} \alpha_s^{CR} P_i^{CR} T_{st} + \sum_{s=1}^{3} \alpha_s^{OS} P_i^{OS} T_{st} + \sum_{s=1}^{3} \alpha_s^{BT} P_i^{BT} T_{st} + v_i + \eta_t + \varepsilon_{it}, \quad (2)$$

where superscripts, *CR*, *OS*, and *BT*, indicate the classroom, on-site, and both training programs, respectively, t is an integer indicating data point with t = 0 meaning the baseline survey and t = 1, 2, 3 meaning the first, second, and third follow-up survey, respectively. The other variables in equation (2), i.e., v_i, η_t, and ε_{it}, stand for unobservable enterprise heterogeneity, a temporary shock common to all enterprises, and an error term, respectively. The definition of the participation dummy, $P^k{}_i$, may be illustrated as follows. For example, $P^{CR}{}_i$ is 1 if enterprise i participated in the classroom training program, and it is 0 otherwise. This definition implies that $P^{CR}{}_i$ = $P^{OS}{}_i$ = $P^{BT}{}_i$ = 1 for enterprises i in Group TT. Similarly, the time dummies, T_{st}'s, are 1 if t = s and 0 otherwise. Under this specification, $\alpha_1{}^{CR}$, $\alpha_2{}^{CR}$, and $\alpha_3{}^{CR}$ represent the impact of participating only in the classroom training that appeared soon after that training, soon after the on-site training, and two years later, respectively. What we call the sustained effect of the classroom training is represented by $\alpha_3{}^{CR}$ while the sustained

17 The ages of entrepreneurs in our data are those as of the baseline survey and hence time-invariant.

effect of both training programs is $\alpha_3^{CR} + \alpha_3^{OS} + \alpha_3^{BT}$. If α^{BT}_3 is positive, the two programs are complements in the sense that the impact of participating in both programs is greater than the sum of the impact of participating only in the classroom training and that of participating only in the on-site training.

We apply two econometric strategies to estimate equation (1). The first is the intention-to-treat analysis, in which the participation variable, P_i, is replaced by the invitation variable, Z_i, which is 1 if enterprise i was invited to at least one of the two training programs. The second is to instrument P_i with Z_i and to estimate the impact on those who comply with the random treatment assignments (Imbens and Angrist, 1994). Because of the low compliance rate of the classroom training, the first approach is expected to result in underestimates of the sustained impact α, and the second approach is expected to have a greater estimate. Similarly, in the estimation (2), we will employ both the intention-to-treat approach and the instrumental variable (IV) approach. In the latter approach, the interaction term $P^k_iT_{st}$ is instrumented with $Z^k_iT_{st}$ for each s and $k = CR$ and BT.[18]

Although we randomized invitation, the sample size is so small that we cannot be sure that ε_i in equation (1) and $v_i + \varepsilon_{it}$ in equation (2) are uncorrelated with Z_i. The fixed effects (FE) model removes the influence of time-invariant heterogeneity, υ_i, but the ANCOVA model, which inserts the mean of *ex ante* values of the dependent variable in vector $\boldsymbol{X_i}$ on the right-hand side, is preferable to the FE model when the dependent variable has low autocorrelation (McKenzie, 2012). We use the ANCOVA in the estimation of (1) and both ANCOVA and FE in the estimation of (2). In the estimation

[18] Since all those firms which were invited to the on-site training program participated in it, $P^{OS}_i = Z^{OS}_i$ and hence P^{OS}_i does not need to be instrumented.

of equations (1) and (2), standard errors are robust to heteroscedasticity and not clustered because data points are few.

These estimators of training impacts should be consistent under the Stable Unit Treatment Value Assumption (SUTVA), an assumption that the outcome variables of non-participants are not affected by participants (Rubin, 1978). The virtual constancy of the *Kaizen* score and WTP of the control group, as shown in columns 4 and 9 of Table 3, is suggestive of the validity of SUTVA, but we cannot be sure. Moreover we do not have any data capturing the degree of knowledge spillovers. Thus, we cannot help but ignore possible spillovers and underestimate the true impacts of the training.[19]

4.4. Econometric Analyses

The estimates of coefficient α in equation (1) are presented in Table 5. Panel A reports the estimates of the intention-to-treat effects of the training on the *Kaizen* score, WTP, and the log of value added.[20] Panels B and C report the estimates of the treatment effects on the treated, which are obtained by applying the instrumental variable method.[21] In Panel B, the sample enterprises are defined as treated if they received a certificate by participating in at least ten days of classroom training and/or received on-site training. In Panel C, the sample enterprises are defined as treated if they participated in at least one day of the classroom training and/or received on-site

19 Bloom, Schankerman et al. (2013) point out that if training participants increase their marketing capabilities and steal business from non-participants, the SUTVA is violated so that the training impact is overestimated. We are not concerned with this possibility, however, because the size of the product market is far greater than the production capacity of the cluster as a whole.

20 The intention-to-treat effect may have more immediate relevance for policy makers than the average effect on participants since the non-participation of some of the invited persons is the rule, not the exception in training programs (McKenzie and Woodruff, 2014).

21 Precisely speaking, the IV estimates capture the differences between the treated and the untreated after the training intervention, not the training effects, when SUTVA is violated.

training. The large values of the first-stage *F*-statistics indicate that the instrumental variable Z_i is valid. These approaches are employed together with the ANCOVA model in which the baseline value of the outcome variable are included in the vector of controls $\boldsymbol{X_i}$ on the right-hand side.

As shown in columns 1 and 4, the training had a positive and significant effect on the *Kaizen* score in both clusters. Participation in at least one of the two training programs increased the number of *Kaizen* practices in use in production two years after the training programs by 1.96 practices in the steel cluster and by 3.49 practices in the knitwear cluster. The short-term training programs like our experimental training program can improve the basic management skills of training participants and such effects can be sustained for two years or longer.

Columns 2 and 5 present the estimated training impacts on WTP. As we did not collect data on WTP in the third follow-up survey, the estimated coefficient are supposed to capture the impacts as of the second follow-up survey. The WTP of the training participants increased by 27% point in the steel cluster and by 43% point in the knitwear cluster. These changes are statistically significant and suggest that a non-negligible number of participants find the training worth three million VND. But there is a caveat to this interpretation: the hypothetical WTP may be merely a proxy for the desire of the participants to please the enumerator. The use of a more convincing measure, such as the willingness to pay for participation in advanced training, is deferred to future research.

Columns 3 and 6 of Table 5 report the estimated impacts on the log of real value added in 2012. Although the coefficient is not significant in the steel cluster, it is marginally significant in the knitwear cluster.[22] The magnitude of the estimated coefficient is large: it suggests that training participation doubled real value added in the knitwear cluster. In Section 4.1, we hypothesized that the training increased value added in the knitwear cluster at least partly because of the reduction in overproduction. To check this, the percentage of output unsold, a measure of dead stock, is regressed on training invitation or participation.[23] As shown in column 7, the coefficient is negative and highly significant, which lends support to the hypothesis.

In Panel C, the magnitude of the estimated impacts are somewhat greater than in Panel B, which suggests that those enterprises which participated in the classroom training for less than ten days benefitted from the training. These enterprises may be called participants without certificate. The narrow definition of participation used in Panel B merges them and those who did not attend the training programs at all into the non-participant group. Under the broader definition used in Panel C, the participants with and without certificate constitute the participant group, and the non-participant group consists of only those who did not attend at all. Compared with the narrow definition, the broader definition will increase the average treatment effect on the treated if the participants without certificate benefitted from the training more than the participants with certificate. Even if the participants without certificate benefitted less than the participants with certificate, one can easily show that the broader definition increases the average treatment effect on the treated if the outcome of the former exceeds sufficiently the outcome of those who did not attend at all. The differences between the estimates in Panels B and C can be interpreted in this way. The

22 When we replace the value added by sales revenue, the coefficients become insignificant. The results are available upon request to the corresponding author.

23 In this regression, the right-hand side does not include the baseline value of the percentage of output sold since the baseline data for this variable is not available.

differences, however, are so small that in the remainder of the paper, we will report only the estimates obtained by using the narrow definition of participation.

The estimates reported in Table 5 are influenced by the correlation between attrition and treatment status, which exists as mentioned in Section 3.2 above. To see the extent of the influence, we have conducted a bounds analysis by applying the method developed by Lee (2009). The major result of this analysis, which is presented in Appendix Table A3 and discussed in Appendix 1, is that the qualitative results shown in Table 5 remain unchanged even after the influence of attrition is controlled for.

We turn now to the estimation of equation (2). Tables 6 and 7 present the FE estimates for the steel cluster and the knitwear cluster, respectively. Appendix 2 discusses the FE estimates obtained by pooling the data from the two clusters. The ANCOVA estimates are not reported here as they are similar to the FE estimates. In these tables, columns 1, 3, and 5 show the intention-to-treat estimates while columns 2, 4, and 6 show the instrumental variable method estimates. The combined effect of the two training programs is reported toward the bottom.

In Table 6, the coefficients on the interaction terms including the “Both training” dummy are positive and significant in columns 1 and 2. These results indicate that the combined impact of the two programs is stronger than the sum of the individual impacts, which suggests the programs are complements. The impacts of the programs on value added, however, are generally insignificant. In Table 7, the two training programs do not seem to be complements, but column 2 suggests that they are substitutes. In columns 5 and 6, the estimated impact of the on-site training on value added is marginally significant. This is probably because that on-site coaching on concrete steps to eliminate wastes was particularly useful even though the same steps were mentioned in the classroom training program as well. The coefficients on the interaction terms including the "Both training" dummy are insignificant but negative, making the total impact of participating in both training programs on value added insignificant. When the data from the two clusters were pooled, the estimates of the training effects are somewhere between the estimates reported in Tables 6 and 7, but because of the increased statistical power, the significance of the estimated sustained effect of the on-site training on log (value added) becomes the 5 percent level (see Appendix 2 and Appendix Table A4).

In columns 7 and 8 of Table 7, the coefficients on the classroom training and the on-site training are negative and highly significant, indicating that each training program reduced dead stock, but the coefficients on the "Both training" dummy is positive because the percentage of output unsold is censored at zero. We also examined what happens if the estimates of training impacts are allowed to vary across enterprises.[24] The qualitative results remain unchanged. An additional finding is that the enterprise owners with low levels of education benefitted more from the training. A possible interpretation of this result is that the less educated business owners can catch up with other owners in terms of management skills if they receive *Kaizen* training even in a short-term program.

5. Conclusion

This study has taken advantage of the randomized design of training intervention and the panel data covering three years to address two questions. Firstly, while the previous studies establish that even short-term training can improve management

24 The estimates of heterogeneous impacts are available upon request to the corresponding author.

practices of small enterprises, the sustainability of improved management practices has been an open question. This paper has found that the *Kaizen* training had favorable effects on practices, and that these effects lasted at least for two years. Secondly, the reason why many enterprise owners and managers do not train themselves and their workers has remained unknown in the literature. Our data suggest that many enterprise owners do not know the importance of learning about management. The results of our analysis indicate that even a classroom training program, which can easily be scaled up, can make participants appreciate the value of learning about management.

In the knitwear cluster, the training participants increased value added by applying the steps to reduce the dead stock of finished products as they learned from the instructors. This effect was not statistically significant immediately after the training but became significant two years later. It seems to take training participants some time to assimilate the knowledge taught in the training program into their production activities and to accumulate working capital to purchase a greater quantity of materials or more expensive materials to expand their businesses probably because most of them were more or less under credit constraints.

In the steel cluster, introductory *Kaizen* activities, such as workshop layout changes, would be prohibitively costly, and unsurprisingly our training programs failed to have significant impacts on their business performance. The contrast between the results in the two study sites points to the importance of designing training contents appropriately.

We have argued that one of the reasons that enterprise owners do not demand training is because they do not know the importance of training. While this argument is speculative and not supported by evidence, it suggests that it is possible to charge the marginal cost of providing advanced training classes and on-site training for those who have come to appreciate the value of learning about management if they are initially given an opportunity to receive basic training probably for free. This hypothesis is testable, but its test is deferred to future research. Another issue for future research is whether management training can be scaled up. This paper attempted to compare the impacts of the classroom and on-site training programs. The results are mixed, however, and the complementarity of the two programs is unclear. To find out what training design is effective, considerable further compilation of studies is warranted.

Appendix 1

As Appendix Table A2 shows, whether or not sample enterprises could be traced in the third follow-up survey depends on treatment status, particularly in the knitwear cluster. To address the issue of possible attrition biases, we employed the bounds analysis method developed by Lee (2009). Appendix Table A3 reports the results for narrowly defined participants.

The 95 percent confidence interval for the *Kaizen* score does not include zero, which indicates that the impact of the training on this variable is still significant at the 5 percent level if attrition is taken into account. The data on WTP are available only up to the second follow-up survey, that is, before attrition occurred, and, hence, bounds analyses are not conducted for this variable. The 95 percent confidence interval for value added is wide and includes zero for both the steel and knitwear samples, but for the knitwear sample, the lower bound is only slightly below zero and the impact is significant at the 10 percent level. For the reduced dead stock, the confidence interval is far away from zero. These results indicate that the results shown in Table 5 remain qualitatively the same if the influence of attrition is controlled for.

Appendix 2

Appendix Table A4 presents the estimates of equation (2) in the pooled data of all the sample enterprises in the two clusters. Generally, the estimated coefficients in Appendix Table A4 lie somewhere between their counterparts in Tables 6 and 7, but the significance levels tend to be higher because of the increased statistical power. An important example is that the estimated sustained effect of the on-site training on log (value added) is insignificant in Table 6, significant at the 10 percent level in Table 7, and significant at the 5 percent level in Appendix Table A4.

References

Arnold, H.J., Feldman, D.C., 1981. Social desirability response bias in self-report choice situations. Academy of Management Journal 24, 377-385.

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, 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.

Bjorvatn, K., Tungodden, B., 2010. Teaching business in Tanzania: Evaluating participation and performance. Journal of European Economic Association 8, 561-570.

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

Bloom, N., Schankerman, M., van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica 81, 1347-1393.

Bloom, N., Van Reenen, J., 2007. Measuring and explaining management practices across firms and countries. Quarterly Journal of Economics 122, 1341-1408.

Bloom, N., Van Reenen, J., 2010. Why do management practices differ across firms and countries? Journal of Economic Perspective 24, 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, 114-137.

Bruhn, M., Karlan, D., Schoar, A., 2010. What capital is missing in developing countries? American Economic Review: Papers and Proceedings 100, 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, 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.

Drexler, A., Fischer, G., Schoar, A., 2014. Keeping it simple: Financial literacy and rules of thumb. American Economic Journal: Applied Economics 6, 1-31.

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, 125-129.

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.

Japan International Cooperation Agency (JICA), 2004. The Study on Artisan Craft Development Plan for Rural Industrialization in the Socialist Republic of Vietnam. Hanoi: JICA.

Karlan, D., Valdivia, M., 2011. Teaching entrepreneurship: Impact of business training of microfinance clients and institutions. Review of Economics and Statistics 93, 510-527.

Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Review of Economic Studies 76, 1071-1102.

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, 458-468.

McKenzie, D., 2012. Beyond baseline and follow-up: The case for more T in experiments. Journal of Development Economics 99, 210-221.

McKenzie, D., Woodruff, C., 2014. What are we learning from business training and entrepreneurship evaluations around the developing world? World Bank Research Observer 29, 48-82.

Mottaleb, K.A., Sonobe, T., 2011. An inquiry into the rapid growth of the garment industry in Bangladesh. Economic Development and Cultural Change 60, 67-89.

Nam, V.H., Sonobe, T., Otsuka, K., 2009. An inquiry into the transformation process of village-based industrial clusters: the case of an iron and steel cluster in northern Vietnam. Journal of Comparative Economics 37, 568-581.

Nam, V.H., Sonobe, T., Otsuka, K., 2010. An inquiry into the development process of village industries: the case of a knitwear cluster in northern Vietnam. Journal of Development Studies 46, 312-330.

Oostendorp, R.H., Trung, T.Q., Tung, N.T., 2009. The changing role of non-farm household enterprises in Vietnam. World Development 37, 632-644.

Rubin, D.B., 1978. Bayesian inference for causal effects: The role of randomization. Annals of Statistics 6, 34-58.

Sonobe, T., Otsuka, K., 2014. Cluster-Based Industrial Development: KAIZEN Management for MSE Growth in Developing Countries. New York: Palgrave Macmillan.

Suzuki, A. Nam, V.H., Sonobe, T., 2014. Willingness to pay for managerial training: a case from the knitwear industry in Northern Vietnam. Journal of Comparative Economics 43, 693-707.

Syverson, C., 2011. What determines productivity? Journal of Economic Literature 49, 326-365.

Tauchmann, T. 2014. Lee (2009) treatment-effect bounds for nonrandom sample selection. Stata Journal 14, 884-894.

Van Biesebroeck, J., 2003. Productivity dynamics with technology choice: An application to automobile assembly. Review of Economic Studies 70, 167-198.

Table 1: Schedule of the data collection and the training programs

	(1) Steel	(2) Knitwear
Baseline survey	2010 June (*N*=153)	2010 July (*N*=159)
Classroom training program	2010 June-July or September	2010 July-August
1st follow-up survey	2010 October (*N*=153)	2010 September (*N*=159)
On-site training program	2010 December-2011 February	2010 December-2011 January
2nd follow-up survey	2011 April (*N*=153)	2011 April (*N*=159)
3rd follow-up survey	2013 January (*N*=128)	2013 January (*N*=146)

Note: *N* stands for the number of sample enterprises. Two model enterprises in each cluster are excluded from the sample.

Table 2: Balance check

	(1) Steel Group TT coefficient	(2) Steel Group TC coefficient	(3) Steel Group CT coefficient	(4) Steel Group CC mean	(5) Steel Overall *F*-test *p*-value	(6) Knitwear Group TT coefficient	(7) Knitwear Group TC coefficient	(8) Knitwear Group CT coefficient	(9) Knitwear Group CC mean	(10) Knitwear Overall *F*-test *p*-value
Panel A										
Owners: Age	2.47	0.76	0.89	43.7	0.58	-0.39	-0.011	-1.89	44.2	0.88
	(1.28)	(0.43)	(0.32)	[45]		(-0.19)	(-0.01)	(-0.73)	[45]	
Male	-0.10	-0.14	-0.071	0.57	0.61	-0.071	0.069	0.086	0.35	0.54
(yes = 1)	(-0.83)	(-1.34)	(-0.39)	[1]		(0.10)	(-0.68)	(0.74)	[0]	
Years of education	-0.36	-0.38	-0.97	7.2	0.82	-0.75	-0.52	0.13	8.5	0.56
	(-0.48)	(-0.61)	(-0.93)	[7]		(-1.26)	(-0.89)	(0.13)	[7.5]	
Business training experience	0.0027	-0.015	0.071	0.03	0.76	0.10	0.085	0.19*	0.06	0.15
(yes = 1)	(0.06)	(-0.49)	(0.71)	[0]		(1.39)	(1.51)	(1.70)	[0]	
Panel B										
Baseline *Kaizen* score	0.20	0.21	0.49	6.1	0.83	-0.17	-0.22	0.64	3.8	0.42
(0-11)	(0.57)	(0.73)	(0.80)	[6]		(-0.64)	(-0.89)	(1.14)	[3]	
Baseline willingness to pay	0.096	0.050	0.17	0.03	0.27	0.17*	0.047	0.26**	0.11	0.09*
(yes=1)	(1.47)	(1.19)	(1.31)	[0]		(1.86)	(0.72)	(2.03)	[0]	
Baseline value added	131.4	-54.5	622.8	1,744	0.80	-275.1	-704.5	30.2	1,437	0.22
(mil. VND)	(0.34)	(-0.14)	(0.86)	[980]		(-0.51)	(-1.41)	(0.04)	[495]	
Baseline sales revenue	5193.3	-558.6	14213.3	26,316	0.48	-305.8	-1616.6	1296.6	4,400	0.14
(mil. VND)	(0.97)	(-0.12)	(1.14)	[16,000]		(-0.26)	(-1.51)	(0.60)	[1,656]	
Baseline employment size	5.82*	-0.67	3.33	19.4	0.20	-4.31	-10.7	9.34	22.4	0.11
	(1.66)	(-0.27)	(0.56)	[17]		(-0.52)	(-1.64)	(0.70)	[8]	
No. enterprises in the Group	32	76	10	35	153	32	57	16	54	159

Notes: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The coefficients and *t*-statistics shown in columns 1 to 3 and 6 to 8 were obtained from the regression of each characteristic on the group dummies with Group CC being the reference. Numbers in parentheses are *t*-statistics based on standard errors robust to heteroskedasticity. Numbers in bracket in columns 4 and 9 are medians. Numbers in columns 5 and 10 are *p*-values from *F*-tests concerning the null hypothesis that all the coefficients on the treatment status dummies are zero. 1 million VND was approximately equivalent to 61 USD for the baseline value added and sales revenue.

Table 3: Deviations in *Kaizen* score, WTP, value added and overproduction between the treatment groups and the control group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Steel	Steel	Steel	Steel	Steel	Knitwear	Knitwear	Knitwear	Knitwear	Knitwear
	Group	Group	Group	Group	Overall	Group	Group	Group	Group	Overall
	TT	TC	CT	CC	*F*-test	TT	TC	CT	CC	*F*-test
	coefficient	coefficient	coefficient	mean	*p*-value	coefficient	coefficient	coefficient	mean	*p*-value
Kaizen score (0-11)										
between the two programs	1.08***	0.46	0.43	6.2	0.03**	1.07***	0.43	0.57	3.9	0.02**
	(3.05)	(1.54)	(0.70)	[6]		(3.06)	(1.50)	(1.00)	[3]	
soon after the on-site training	2.33***	0.41	0.90*	6.2	0.00***	3.85***	0.37	3.48***	4.0	0.00***
	(5.90)	(1.37)	(1.71)	[7]		(13.11)	(1.28)	(10.93)	[4]	
two years after the programs	2.68***	0.31	0.92*	6.4	0.00***	4.33***	1.32***	4.17***	3.9	0.00***
	(9.24)	(1.39)	(1.90)	[6]		(17.80)	(3.93)	(15.75)	[3.5]	
Willingness to pay (=1 if yes)										
between the two programs	0.19**	0.15**	0.14	0.06	0.04**	0.56***	0.26***	0.25*	0.13	0.00***
	(2.21)	(2.48)	(1.06)	[0]		(5.87)	(3.20)	(1.87)	[0]	
soon after the on-site training	0.32***	0.15**	0.14	0.06	0.00***	0.65***	0.25***	0.37***	0.13	0.00***
	(3.33)	(2.48)	(1.06)	[0]		(7.40)	(3.15)	(2.72)	[0]	
Real value added (mil. VND in 2008)										
in 2010	1538.2*	154.6	3329.9	2,950	0.11	-0.59	-285.6	223.5	1,150	0.55
	(1.94)	(0.26)	(1.57)	[1,531]		(-0.00)	(-1.01)	(0.39)	[481]	
in 2012	177.2	141.4	576.2	1,968	0.97	606.0	-90.3	1974.7	657	0.10
	(0.28)	(0.19)	(0.49)	[847]		(1.57)	(-0.46)	(1.51)	[244]	
% output unsold by the end of year										
in 2012						-7.10***	-5.08**	-7.20***	7.20	0.00***
						(-3.46)	(-2.33)	(-3.51)	[0]	

Notes: ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The coefficients and *t*-statistics shown in columns 1 to 3 and 6 to 8 were obtained from the regression of each variable in the most left-hand column on the group dummies with Group CC being the reference. Numbers in parentheses are *t*-statistics based on standard errors robust to heteroskedasticity. Numbers in bracket in columns 4 and 9 are medians. Numbers in columns 5 and 10 are *p*-values from *F*-tests concerning the null hypothesis that all the coefficients on the treatment status dummies are zero. The sample sizes are the same as shown in Table 2 except for those for *Kaizen* score and real value added measured two years after the programs and those for percentage of output unsold because the 25 steel and 12 knitwear enterprises did not operate in 2012.

Table 4: *Ex ante* correlates of *Kaizen* score, WTP, value added and participation decision to the classroom training

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline *Kaizen* score	Baseline WTP	Baseline log (value added)	Baseline log (value added)	*y* = 1 if participated	*y* = 1 if participated
	OLS	OLS	OLS	OLS	Probit	Probit
Steel cluster						
Age	0.078	-0.0030	0.17**	0.15*	0.050	0.064
	(0.56)	(-0.11)	(2.31)	(1.93)	(0.27)	(0.34)
Age squared (/100)	-0.11	0.0088	-0.19**	-0.16*	-0.060	-0.079
	(-0.69)	(0.26)	(-2.09)	(-1.66)	(-0.30)	(-0.38)
Male (yes = 1)	-0.45*	-0.019	-0.27**	-0.16	-0.32	-0.36
	(-1.90)	(-0.34)	(-2.00)	(-1.38)	(-1.15)	(-1.26)
Years of education	0.11**	0.029***	0.10***	0.077***	0.035	0.040
	(2.18)	(2.78)	(4.36)	(3.53)	(0.63)	(0.73)
Business training experience	0.50	-0.10**	0.51	0.39		
(yes = 1)	(0.58)	(-2.14)	(1.14)	(1.50)		
Baseline *Kaizen* score				0.24***		-0.070
				(6.02)		(-0.72)
Baseline WTP					1.32***	1.40***
					(2.77)	(2.76)
No. observations	153	153	153	153	106	106
Knitwear cluster						
Age	0.14*	0.018	0.13	0.074	0.065	0.058
	(1.68)	(0.65)	(0.38)	(0.22)	(0.38)	(0.34)
Age squared (/100)	-0.12	-0.025	-0.15	-0.10	-0.12	-0.12
	(-1.12)	(-0.76)	(-0.37)	(-0.26)	(-0.62)	(-0.60)
Male (yes = 1)	0.44**	0.069	-0.62	-0.78	0.61*	0.58*
	(2.34)	(1.04)	(-0.82)	(-1.00)	(1.90)	(1.80)
Years of education	0.22***	0.0091	0.014	-0.066	-0.0065	-0.025
	(4.79)	(0.77)	(0.08)	(-0.39)	(-0.11)	(-0.39)
Business training experience	0.72*	0.049	1.95**	1.69*	0.21	0.13
(yes = 1)	(1.93)	(0.49)	(2.27)	(1.86)	(0.54)	(0.34)
Baseline *Kaizen* score				0.36		0.10
				(1.44)		(0.83)
Baseline WTP					1.31***	1.27***
					(2.97)	(2.88)
No. observations	159	159	159	159	89	89

Notes: ***, **, and * indicate the 1%, 5%, and 10 percent level of statistical significance, respectively. Numbers in parentheses are *t*-statistics robust to heteroskedasticity. Baseline value added for the steel cluster is annual value added in 2008 while that for the knitwear cluster is the average of annual value added in 2008 and 2009. In columns 5 and 6, the sub-sample of Groups TT and TC is used. The past business training experience dummy is omitted in columns 5 and 6 in the steel cluster because all the enterprises in the sub-sample with the training experience participated in the classroom training.

Table 5: Sustained impacts of the training programs

	(1) Steel *Kaizen* score	(2) Steel WTP	(3) Steel log (value added)	(4) Knitwear *Kaizen* score	(5) Knitwear WTP	(6) Knitwear log (value added)	(7) Knitwear % output sold
A. *Intention-to-treat effects*							
Coefficient on Z_i (i.e., invitation to at least one of the two training programs)	1.09***	0.14***	0.17	2.56***	0.30***	0.78*	-6.25***
	(5.81)	(3.00)	(0.25)	(10.46)	(5.32)	(1.76)	(-2.81)
B. *Treatment effects on the narrowly defined participants*							
Coefficient on P_i (i.e., participation in at least one program) instrumented with Z_i	1.96***	0.27***	0.30	3.49***	0.43***	1.06*	-8.56***
	(7.62)	(3.35)	(0.26)	(16.48)	(5.49)	(1.81)	(-2.87)
First stage *F*-statistics	29.3	34.9	29.2	44.9	132.3	44.5	49.0
C. *Treatment effects on the broadly defined participants (including those without certificate)*							
Coefficient on P_i (i.e., participation in at least one program) instrumented with Z_i	2.02***	0.28***	0.31	3.76***	0.46***	1.14*	-9.23***
	(7.52)	(3.37)	(0.26)	(13.74)	(5.55)	(1.81)	(-2.85)
First stage *F*-statistics	31.6	33.1	33.1	93.5	103.0	95.6	91.7
No. of observations (= No. of enterprises)	128	153	128	146	158	146	146

Notes: For the intention-to-treat effects, the reported is the coefficients of the dummy variable taking one if the enterprise was invited to any of the training programs. For the treatment effects on the treated, the reported is the coefficients of the dummy variable taking one if the enterprise participated in any of the training programs, instrumented by the dummy variable taking one if the enterprise was invited to any of the training programs. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Numbers in parentheses are *t*-statistics based on standard errors robust to heteroskedasticity. The estimation uses the cross-section data of *Kaizen* score collected two years after the training programs for, that of WTP collected soon after the training programs, that of value added for 2012, and that of percentage of output unsold for 2012. In all estimation, owners' age, its squared, male dummy, years of education and business training experience are controlled although the coefficients are not reported. For the estimation of *Kaizen* score, WTP and value added, the baseline value of each outcome is also controlled while the coefficients are not reported.

Table 6: Impacts of the training programs (Steel cluster)

	(1)	(2)	(3)	(4)	(5)	(6)
	Kaizen score		WTP		log (value added)	
	FE	FE-IV	FE	FE-IV	FE	FE-IV
Classroom training	0.25**	0.89***	0.10*	0.37***		
* After CR	(2.53)	(2.66)	(1.73)	(2.71)		
Classroom training	0.21*	0.75**	0.10*	0.37***	-0.32	-1.16
* After OS	(1.89)	(2.20)	(1.73)	(2.71)	(-1.10)	(-0.82)
Classroom training	0.40	1.40**			0.21	0.68
* 2yrsAfter	(1.55)	(2.19)			(0.34)	(0.34)
On-site training	-0.057	-0.057	-0.029	-0.029		
* After CR	(-1.23)	(-0.27)	(-0.82)	(-1.11)		
On-site training	0.41	0.41*	-0.029	-0.029	0.18	0.18
* After OS	(1.35)	(1.72)	(-0.82)	(-1.11)	(1.08)	(0.63)
On-site training	0.88	0.88**			0.79	0.79
* 2yrsAfter	(1.55)	(2.08)			(1.48)	(1.63)
Both training	0.69***	0.61	0.022	-0.17		
* After CR	(3.49)	(1.14)	(0.23)	(-1.04)		
Both training	1.51***	2.00***	0.15	0.027	0.41	1.30
* After OS	(3.25)	(3.55)	(1.31)	(0.16)	(1.19)	(0.86)
Both training	1.47**	1.62*			-0.71	-1.51
* 2yrsAfter	(2.11)	(1.65)			(-0.89)	(-0.70)
After CR	0.057	0.057	0.029	0.029		
	(1.23)	(0.84)	(0.82)	(1.11)		
After OS	0.086	0.086	0.029	0.029	0.49***	0.49**
	(1.53)	(1.26)	(0.82)	(1.11)	(7.07)	(2.06)
2yrsAfter	0.12	0.12			-0.69	-0.69
	(0.60)	(0.78)			(-1.44)	(-1.60)
Total impact of both	2.13***	3.16***	0.22**	0.37***	0.26**	0.31
training (After OS)	(6.23)	(10.96)	(2.18)	(3.49)	(2.32)	(0.61)
Total impact of both	2.75***	3.91***			0. 29	-0.03
training (2yrsAfter)	(6.66)	(8.46)			(0.43)	(-0.04)
Cragg-Donald *F*-statistic		8.1		9.0		8.2
No. observations	586	586	459	459	434	434
No. enterprises	153	153	153	153	153	153

Notes: ***, **, and * indicate the 1%, 5%, and 10 percent level of statistical significance, respectively. Numbers in parentheses are *t*-statistics based on standard errors robust to heteroskedasticity. The interaction terms including the classroom participation dummy or the both programs participation dummy are instrumented with the interaction between the corresponding invitation status dummies and the time dummies in columns 2, 4, and 6. Since the WTP data were not collected at the third follow-up survey (two years after the training programs), the 2yrsAfter dummy and its interaction with the participation status dummies are not included in the empirical model explaining WTP (see columns 3 and 4). In columns 5 and 6, the After OS dummy and the 2yrsAfter dummy were redefined: the former takes one if the year is 2010 and zero otherwise, and the latter takes one if the year is 2012 and zero otherwise.

Table 7: Impacts of the training programs (Knitwear cluster)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Kaizen score		WTP		log (value added)		% output unsold	
	FE	FE-IV	FE	FE-IV	FE	FE-IV	OLS	IV
Classroom training	0.65***	1.23***	0.21***	0.40***				
* After CR	(3.91)	(5.96)	(2.88)	(4.17)				
Classroom training	0.61***	1.15***	0.21***	0.40***	0.94	1.79		
* After OS	(3.59)	(5.59)	(2.87)	(4.06)	(1.22)	(1.26)		
Classroom training	1.41***	2.74***			1.19	2.31	-5.25**	-10.2**
* 2yrsAfter	(4.97)	(10.70)			(1.38)	(1.44)	(-2.34)	(-2.38)
On-site training	-0.074	-0.074	-0.019	-0.019				
* After CR	(-1.42)	(-0.24)	(-0.81)	(-0.37)				
On-site training	2.85***	2.85***	0.11	0.11	1.03	1.03		
* After OS	(6.77)	(9.47)	(1.01)	(1.37)	(0.83)	(0.90)		
On-site training	3.36***	3.36***			2.18*	2.18*	-7.45***	-7.39***
* 2yrsAfter	(6.78)	(9.55)			(1.71)	(1.86)	(-3.28)	(-3.40)
Both training	0.67**	0.68	0.20	0.19				
* After CR	(2.15)	(1.20)	(1.51)	(1.24)				
Both training	0.58	0.58	0.17	0.15	-0.84	-1.65		
* After OS	(1.11)	(1.03)	(0.99)	(0.82)	(-0.52)	(-0.69)		
Both training	-0.48	-1.40**			-1.66	-2.96	5.10**	10.1**
* 2yrsAfter	(-0.79)	(-2.28)			(-0.99)	(-1.20)	(2.31)	(2.38)
After CR	0.074	0.074	0.019	0.019				
	(1.42)	(1.41)	(0.81)	(1.11)				
After OS	0.15**	0.15***	0.019	0.019	0.75	0.75		
	(2.28)	(2.79)	(0.81)	(1.10)	(1.57)	(1.60)		
2yrsAfter	0.26***	0.26***			0.048	0.048		
	(2.73)	(3.55)			(0.08)	(0.08)		
Total impact of both	4.04***	4.57***	0.48***	0.65***	1.12	1.17		
training (After OS)	(14.41)	(14.15)	(4.29)	(5.93)	(1.14)	(0.87)		
Total impact of both	4.29***	4.71***			1.72	1.53	-7.60***	-7.48***
training (2yrsAfter)	(16.79)	(15.02)			(1.57)	(1.11)	(-3.19)	(-3.23)
Cragg-Donald *F*-statistic		36.2		41.4		35.8		33.4
No. observations	622	622	476	476	464	464	146	146
No. enterprises	159	159	159	159	159	159	146	146

Notes: ***, **, and * indicate the 1%, 5%, and 10 percent level of statistical significance, respectively. Numbers in parentheses are *t*-statistics based on standard errors robust to heteroskedasticity. The interaction terms including the classroom participation dummy or the both programs participation dummy are instrumented with the interaction between the corresponding invitation status dummies and the time dummies in columns 2, 4, 6 and 8. Since the WTP data were not collected at the third follow-up survey, the 2yrsAfter dummy and its interaction with the participation status dummies are not included in the empirical model explaining WTP (see columns 3 and 4). In columns 5 and 6, the After OS dummy and the 2yrsAfter dummy were redefined: the former takes one if the year is 2010 and zero otherwise, and the latter takes one if the year is 2012 and zero otherwise. Columns 7 and 8 use the cross-section data in 2012. Although their coefficients are not reported, owners' age, its squared, male dummy, years of education and business training experience are controlled in columns 7 and 8.

Appendix Table A1: Components of the *Kaizen* score

Evaluation based on the enumerators' visual inspections
The enterprise has a designated area for each production/activity within the workshop.
The enterprise has a fixed place where major tools are stored.
The storage of tools is put in order by kind.
The enterprise has a fixed place where raw materials are stored.
The raw materials are stored separately from the scrap.
The work flow line is determined.
The defectives of raw materials and finished products are clearly segregated from the good ones.
Evaluation based on the owners' responses
The scraps are removed and the floor is cleaned every day.
The workers maintain machines every day.
The enterprise holds meeting in which all workers participate.
The proprietor knows how long each production process takes.

Appendix Table A2: Predicting attrition (y =1 if included in the third follow-up survey)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Steel	Steel	Steel	Steel	Knitwear	Knitwear	Knitwear	Knitwear
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Any training	0.011	0.021			0.11**	0.14**		
	(0.15)	(0.16)			(2.25)	(2.27)		
Invited to classroom training			-0.0066	-0.013			0.083	0.15
			(-0.08)	(-0.05)			(1.58)	(1.55)
Invited to on-site training			-0.25	-0.25			0.18***	0.17***
			(-1.53)	(-1.55)			(3.15)	(3.23)
Invited to both programs			0.40**	0.65*			-0.14**	-0.24**
			(2.30)	(1.71)			(-2.07)	(-2.02)
Age	-0.034	-0.034	-0.041	-0.047	0.0090	0.0066	0.0069	0.0088
	(-0.95)	(-0.98)	(-1.13)	(-1.22)	(0.37)	(0.27)	(0.28)	(0.36)
Age squared (/100)	0.041	0.041	0.048	0.055	-0.0099	-0.0065	-0.0074	-0.0092
	(1.09)	(1.12)	(1.25)	(1.34)	(-0.37)	(-0.24)	(-0.27)	(-0.34)
Male	0.025	0.025	0.025	0.063	-0.023	-0.029	-0.022	-0.029
(yes = 1)	(0.40)	(0.41)	(0.42)	(0.92)	(-0.43)	(-0.53)	(-0.41)	(-0.54)
Years of education	-0.018	-0.018	-0.021	-0.027*	-0.013	-0.013	-0.013	-0.013
	(-1.36)	(-1.40)	(-1.64)	(-1.92)	(-1.02)	(-1.04)	(-1.06)	(-1.04)
Business training experience	0.19***	0.18**	0.23***	0.16	-0.14	-0.13	-0.15	-0.13
(yes = 1)	(3.30)	(2.21)	(2.69)	(1.20)	(-1.49)	(-1.46)	(-1.56)	(-1.36)
Total impact of both			0.14*	0.39**			0.12**	0.09
training			(1.85)	(2.51)			(2.09)	(1.28)
No. observations	153	153	153	153	159	159	159	159
No. attrition	25	25	25	25	12	12	12	12

Notes: ***, **, and * indicate the 1%, 5%, and 10 percent level of statistical significance, respectively. Numbers in parentheses are *t*-statistics based on standard errors robust to heteroskedasticity. The any training participation dummy is instrumented with any training invitation dummy. The classroom participation dummy and the both programs participation dummy are instrumented with the classroom invitation dummies and the both programs invitation dummy.

Appendix Table A3: Lee's (2009) 95 percent confidence intervals for the sustained impact estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Steel		Knitwear		Both clusters pooled	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Kaizen score	0.33	1.72	1.81	3.78	1.49	2.97
log (value added)	-1.09	2.80	-0.01	2.00	-0.08	1.66
% output unsold			-10.7	-2.26		
No. obs. with non-missing values	128		146		274	

Notes: The bounds are calculated by "leebounds" command in STATA developed by Tauchmann (2014). The computed bounds are for the sustained impact of being invited to any of the training programs (corresponding to Panel A of Table 5).

Appendix Table A4: Impacts of the training programs (Two clusters pooled)

	(1)	(2)	(3)	(4)	(5)	(6)
	Kaizen score		WTP		log (value added)	
	FE	FE-IV	FE	FE-IV	FE	FE-IV
Classroom training	0.41***	1.08***	0.15***	0.39***		
* After CR	(4.48)	(5.88)	(3.35)	(5.10)		
Classroom training	0.37***	0.96***	0.15***	0.39***	0.17	0.45
* After OS	(3.71)	(5.18)	(3.35)	(5.03)	(0.41)	(0.40)
Classroom training	0.83***	2.15***			0.52	1.34
* 2yrsAfter	(4.46)	(7.42)			(0.97)	(1.02)
On-site training	-0.067*	-0.067	-0.022	-0.022		
* After CR	(-1.86)	(-0.24)	(-1.16)	(-0.67)		
On-site training	1.91***	1.91***	0.054	0.054	0.70	0.70
* After OS	(4.82)	(6.66)	(0.81)	(1.09)	(0.92)	(0.98)
On-site training	2.61***	2.61***			1.83**	1.83**
* 2yrsAfter	(5.50)	(7.63)			(2.07)	(2.22)
Both training	0.71***	0.64	0.12	0.012		
* After CR	(3.88)	(1.23)	(1.41)	(0.11)		
Both training	0.80*	0.81	0.15	0.062	-0.21	0.45
* After OS	(1.65)	(1.53)	(1.36)	(0.47)	(-0.23)	(0.40)
Both training	0.062	-0.78			-1.43	1.34
* 2yrsAfter	(0.11)	(-1.23)			(-1.33)	(1.02)
After CR	0.067*	0.067	0.022	0.022		
	(1.86)	(1.62)	(1.16)	(1.57)		
After OS	0.12***	0.12***	0.023	0.023	0.65**	0.65**
	(2.75)	(2.95)	(1.16)	(1.56)	(2.23)	(2.15)
2yrsAfter	0.21**	0.21***			-0.24	-0.24
	(2.17)	(2.79)			(-0.57)	(-0.61)
Total impact of both	3.07***	3.69***	0.35***	0.51***	0 .66	0 .64
training (After OS)	(11.72)	(12.73)	(4.56)	(6.22)	(1.27)	(0.80)
Total impact of both	3.51***	3.97***			0.92	0.46
training (2yrsAfter)	(13.16)	(12.62)			(1.37)	(0.50)
No. observations	1208	1208	935	935	898	898
No. enterprises	312	312	312	312	312	312
Cragg-Donald *F*-statistic		36.0		38.1		36.0

Notes: ***, **, and * indicate the 1%, 5%, and 10 percent level of statistical significance, respectively. Numbers in parentheses are *t*-statistics based on standard errors robust to heteroskedasticity. The interaction terms including the classroom participation dummy or the both programs participation dummy are instrumented with the interaction between the corresponding invitation status dummies and the time dummies in columns 2, 4, and 6. Since the WTP data were not collected at the third follow-up survey, the 2yrsAfter dummy and its interaction with the participation status dummies are not included in the empirical model explaining WTP (see columns 3 and 4). In columns 5 and 6, the After OS dummy and the 2yrsAfter dummy were redefined: the former takes one if the year is 2010 and zero otherwise, and the latter takes one if the year is 2012 and zero otherwise.