

# Improving the Productivity and Earnings of Women Owned and/or Managed Enterprises in Developing Countries: What Works? \*

Christopher Woodruff, *University of Warwick*  
David McKenzie, *World Bank*

7 December 2012

## **Abstract**

Business training programs are a popular policy option for improving the performance of enterprises owned by women around the world. The last few years have seen rapid growth in the number of rigorous evaluations of these programs in developing countries. We undertake a critical review of these studies with the goals of synthesizing the lessons emerging from them, understanding their drawbacks, and identifying areas in which more work is needed. We focus on studies which involve a randomized control group and those which provide a credible non-experimental design. We find that there is substantial heterogeneity in the length and content of training programs, and in the types of firms participating in them. Most studies find that existing firm owners implement some of the practices taught in training, but the magnitudes of these improvements in practices are often relatively modest. But few studies find significant impacts on profits or sales, although a couple of the studies with more statistical power have found impacts on these outcomes.

Many evaluations suffer from low statistical power, measure impacts only within the first year after training has finished, and experience problems with survey attrition and measuring firm profits and revenues. With these limitations in mind, the existing literature indicates that over the short time horizons analyzed, there are relatively modest impacts of training on survivorship of existing firms, and stronger evidence that training programs have helped prospective owners launch new businesses more quickly. Some studies have also found benefits to microfinance organizations of offering this training. To date there is little evidence to help guide policymakers as to whether any impacts found come from trained firms competing away sales from other businesses versus through productivity improvements, and little evidence to guide the development of the provision of training at market prices. We conclude by summarizing some directions for future studies.

*Keywords:* Business training; Consulting; Randomized experiments; Firm Productivity.

*JEL codes:* O12, J16, L26, M53.

---

\* This report was prepared for the United Nations Foundation. The draft borrows (very) heavily on the draft of “What are we learning from business training and entrepreneurship evaluations around the developing world?” authored jointly by David McKenzie and Christopher Woodruff.

## Introduction

Walk into a typical micro or small business in a developing country and spend a few minutes talking with the owner and it is often clear that they are not implementing many of the business practices standard in well-run small businesses in developed countries. Records are often not kept, and household and business finances are mixed together. Marketing efforts are sporadic and rudimentary. Some inventory sits on shelves for years at a time, while more popular items frequently are out of stock. Few owners have financial targets or goals that they regularly monitor and act to achieve. The picture is not much better in medium and some large firms, with few firms using quality control systems, rewarding workers with performance-based pay, or using many other practices typical in well managed firms in developed countries.

Small wonder, then, that business training is one of the most common forms of active support provided to small firms around the world. In addition to programs provided by national governments and microfinance organizations, there are a number of programs offered in many countries around the world. The International Labor Organization (ILO)'s Start and Improve Your Business (SIYB) program started in 1977 now claims over 4.5 million trainees in over 95 countries. Other widely used programs include the GTZ/CEFE program, the UNCTAD/EMPRETEC program, business plan competitions and training run by Technoserve, content for microfinance clients developed by Freedom from Hunger, and the IFC's Business Edge and SME Toolkit programs.

However, until recently there has been very little rigorous evidence on the impact of these programs. An overview of impact evaluations in finance and private sector development written in 2009 found very little work on business training (McKenzie, 2010). The last three years have seen a rapid increase in attention being given to the idea that "managerial capital" or poor management is a constraint to production in developing countries (Bruhn et al, 2010; Bloom and Van Reenen, 2010), and the emergence of a number of impact evaluations of business training programs. This paper provides a critical overview of what we are learning from such studies, and attempts to draw out the emerging lessons for both policy and the next generation of research.

This review focuses on training programs aimed at female-owned and managed businesses in developing countries. Sometimes we reference programs focusing on male-owned enterprises for comparison purposes. Regardless of gender, an issue which is important in interpreting the results of training programs is the selection of the sample of trainees. This is likely to be especially important for women. Studies of returns to capital, for example, find large differences in outcomes between randomly selected samples of women in Sri Lanka (de Mel et al 2009) and Ghana (Fafchamps et al 2011). As discussed in Fafchamps et al, the difference in outcomes for women are driven by differences in the characteristics of women in the two samples. With regards to training, the target groups for programs are often microfinance clients, but are sometimes applicants to business plan competitions and sometimes random samples. In addition to sample selection, context – and especially norms with respect to female labor force participation – may matter. These factors, combined with the modest number of credible studies which are available, make drawing overall lessons challenging.

We used a variety of methods to identify recent studies which examine the impacts of business training in developing countries. These include an Econlit search for published studies, Google Scholar searches of papers which cite these published studies or other working papers, our contacts with scholars working in this field, input from recent training program inventory exercises, and knowledge of papers presented in recent seminars or conferences. We restrict attention to papers which have a clear impact evaluation design that attempts to deal with selection on both observable and unobservable characteristics of business owners, and which focuses on enterprise management rather than focusing solely on technical or vocational training. This yields 11 randomized experiments of business training, 3 regression discontinuity studies, and 1 experiment which focus on individualized consulting services.

We start by assessing how comparable these different programs are in terms of what training is provided and who takes part in the training. There is considerable heterogeneity across studies in both the program participants and in the length and content of the training provided – although there are a number of core topics that are covered in most training programs. We next discuss a number of challenges these studies have faced in measuring impact. In particular, we find most of the existing studies have very low power for detecting impacts on business profits or revenues. This lack of power results from reliance on a single follow-up survey to measure impacts, on relatively small sample sizes of very heterogeneous firms, and from not all firms who are offered training taking it up, which dilutes the impacts further. In addition, many existing studies look at impacts only within a year of the training, which may be too short to detect some of the impacts. Many also experience problems with attrition, selective survival and / or start-up, and with some firm owners being reluctant to share information on profits and revenues. A final concern is that training changes measurement of outcomes, even if it doesn't change the outcomes themselves. We discuss how several studies have attempted to show their results are robust to the reporting issues.

With these issues in mind, we assess the impacts of different training programs. Relatively few studies have looked at impacts on survivorship of existing businesses. Among programs aimed at female-owned businesses, most find either zero, or slightly negative, impacts on survivorship. (By comparison, there is some weak evidence for a positive survivorship effect for male-owned businesses in a couple of studies.) Stronger results have been found among programs which look at the impacts of training programs on new business start-ups. All the training programs studied which have content specifically for people wanting to start a new business have found training helps firms start-up, although there is some evidence that this may merely speed up the entry of firms who were going to enter anyway, and potentially change the selection of who enters.

Almost all training programs find that treated firms start implementing some of the business practices taught by the training. However, the magnitude of the impacts is rather small in many cases, with a typical change being 0.1 or 0.2 standard deviations, or 5 to 10 percentage points. The combination of relatively small changes in business practices and low statistical power means that few studies find impacts on sales or profitability, although the studies with most power have found some positive short-term effects. Studies which work with microfinance clients have also looked at microfinance institution outcomes, and in some cases found changes in the likelihood of clients being retained or in the characteristics of who applies for loans. We also examine studies which look at the impacts of

individualized consulting being provided to larger firms, with evidence that such services can improve the performance of even firms with multiple plants and over 200 workers.

Most of the studies we review are randomized control trials. Given issues related to selection – and often selection at multiple levels – RCTs will generally provide the most credible evidence. However given heterogeneity in samples, in regions and in training content, conducting enough large scale RCTs to properly assess which types of training are effective for which enterprise owners in which regions may take considerable time and resources to accomplish. We therefore discuss a few non-experimental studies, assessing in particular the credibility of the manner in which they deal with potential selection bias. Though we may always be more confident with randomized designs, we would leave open the question of whether funders interested in learning about the effectiveness of training may want to consider a mixture of RCTs and non-experimental studies, where the latter are used to assess the heterogeneity in outcomes.

Before concluding we then discuss several important issues that existing work provides very little evidence about, but which are crucial for developing policy recommendations. The list is long and includes whether or not any gains from training are long-lasting, and whether they come from competing away sales from other firms or through other channels. There are also issues of heterogeneity of training effects, what the market failures are that might prevent firms from investing in training themselves if it is really beneficial, and what the role of different types of training content is. We then conclude with recommendations for future work in this area.

## **2. What does a typical business training program involve?**

Attempts to measure “the” impact of “business training” face multiple challenges, which complicate comparisons across studies. The first challenge is that business training varies in what is offered and how it is offered across different locations and organizations, with these differences in content likely to be important and to induce much more heterogeneity into the treatment of business training than there is in several other firm interventions such as access to credit or grants. A second challenge is that the impact of training is likely to differ depending on who it is provided to, so that even if we compare the same training content in different locations, differences in the characteristics of the individuals receiving the training may result in different measured impacts. In practice, studies vary in both how they select individuals to participate in them, and what content they offer to those who are selected, so that overall impacts include both heterogeneity in individual impacts and heterogeneity in training content. It is therefore important to examine the heterogeneity in who participates and in what is offered before making comparisons among studies.

### *Who takes part in business training experiments?*

Table 1 summarizes key characteristics of the participants in recent business training evaluation studies. The most common modality for training amongst these studies is for it to be offered by microfinance organizations or banks to their clients. This is particularly common for training offered to female

microenterprise owners, given that many microfinance clients focus on women. In some cases training is offered to all clients as part of their regular group meetings (e.g. Karlan and Valdivia, 2011, Berge et al, 2011) while in others the bank or microfinance organization offers this as an additional service that clients can choose to apply for (e.g. Bruhn and Zia, 2012; Valdivia, 2011). A second strategy has been to choose firms in a particular industrial cluster in a specific industry, and then offer training to them (e.g. Sonobe et al, 2011). A third strategy is to have individuals apply to take part in the training as part of a competition, as is done by Technoserve (Klinger and Schündeln, 2011), or to have students apply to take part in an entrepreneurship course (Premand et al, 2012). All of these approaches result in a selected sample of firms, who may differ from the general population. A final approach, to date used only by de Mel et al. (2012) and Calderon et al (2012), is to conduct a representative sample of the microenterprise population of interest and then offer the training to a random sample from this population.

As a result of these approaches to selecting participants, most evaluations focus on existing businesses. Exceptions include studies where many of the microfinance clients are borrowing or saving for household purposes but do not necessarily have an enterprise (e.g. Field et al, 2010), and studies which are based on competitions or training of new businesses (Klinger and Schündeln, 2011; Premand et al, 2012; de Mel et al, 2012). It is noticeable that almost all the evaluations to date have focused on urban clients, which likely reflects the greater density of businesses and training providers in urban areas. Exceptions are Giné and Mansuri (2011) who work in rural Pakistan, and half of the sample of Karlan and Valdivia (2011), which comes from rural Peru.

The average age of a participant in a typical study is 35 to 45 years old, with only a couple of studies focused on young entrepreneurs (Bruhn and Zia, 2012; Premand et al, 2012). Some studies focus entirely on female business owners, others on entirely male owners, while relatively few have sufficient numbers of both gender to compare impacts separately by gender. Finally, there is substantial heterogeneity in the education levels of the participants, ranging from as low as 2.5 years of schooling for females and 5.7 years for males in the rural Pakistan study of Giné and Mansuri (2011) and up to all university-educated in the course of Premand et al. (2012).

Table 2 shows the degree of heterogeneity across studies in the baseline size of the participating firms for studies that include existing firms. The smallest firms are subsistence firms run by women in Giné and Mansuri (2011) and de Mel et al. (2012), where 95 percent of the firms have no paid employees, and average monthly revenues are only US\$80-100 at market exchange rates, with profits of approximately one dollar per day. Most of the rest of the studies also focus on microenterprises, albeit ones with slightly larger revenues and potentially one or two employees. The main exceptions are the firms chosen from industrial clusters (Mano et al, 2011 and Sonobe et al, 2012), where the firms are SMEs with 5 to 50 workers, and monthly revenues which are \$5000 or more (and in some cases over \$100,000).

#### *Training delivery and costs*

All of the training courses studied here are classroom-based training delivered to groups of individuals, although several of the studies supplement this with additional treatments of one-on-one follow-ups,

which we will discuss later. Table 3 provides key characteristics of the training delivery in the different studies. A first point to note is that many of the studies are testing content that has been modified or developed specifically for the study of interest, rather than content that has been taught for years. This may matter if it takes a while teaching particular content for it to become adapted to local context and for instructors to become familiar with it. The instructors in most studies are experienced training professionals, although in a few studies, it has been the microfinance officers themselves who have been trained to deliver content. Berge et al. (2012) compare the same training content taught by microfinance staff to that taught by professional trainers, and find that attendance rates fall more for the microfinance staff taught course, subjective evaluations of course quality are lower, and understanding of the concept of profits in tests is lower for those taught by the microfinance staff. As a consequence of differences in the newness of the content and in who provides it, the quality of training is likely to vary across studies.

The length of the training course also varies substantially across studies. The shortest courses are two days or two half-days (Bruhn and Zia, 2012; Field et al, 2010), while other courses are for one week or more full-time (e.g. de Mel et al, 2012; Sonobe et al, 2011). In most cases the training is concentrated over a relatively short period of time, but in some of the cases where training takes place in microfinance group meetings, it can be spread over many months, with as little as half an hour per month over many months (e.g. Karlan and Valdivia, 2011). Longer courses allow more content to be taught, but are more costly, and also require business owners to be away for their businesses for more time.

In all these experiments training has been offered for free, and in addition, some studies have also provided small supplements for travel or food, or offered the prospect of grants as an additional incentive. For example, Bruhn and Zia (2012) paid clients approximately \$35 for participating in their two day course to compensate for the opportunity cost of time, while de Mel et al. (2012) paid a travel allowance of slightly less than \$4 per day, but also the incentive that half of those completing training would receive a grant of approximately \$130. The actual costs per person trained range from as little as \$21 in Drexler et al. (2011) where training was conducted by local instructors once per week over five or six weeks in local schools, up to more than \$400 per firm reported in Sonobe et al. (2011), where instructor costs and venue rental costs per person for 15 days were relatively high. One argument for subsidizing costs is that because many business owners have little perception of how badly managed their firms are, and training is a new and unproven concept to them with uncertain payoffs, even those who are not liquidity constrained may be reluctant to pay. Indeed, two studies find some evidence to support the idea that individuals who are least interested have potentially the most to gain from training (Bjorvatn and Tungodden, 2010; Karlan and Valdivia, 2011). We will return to discussing market failures and subsidies later in the paper, after discussing training impacts.

Despite training being offered for free, take-up of training is far from universal. The average take-up rate across the different studies for individuals offered training is approximately 65 percent. Low take-up reduces the power of an experiment, with 2.4 times the sample size needed to detect a given treatment effect when take-up is 65% than when it is 100%. One would expect take-up rates to be highest when training takes place in the context of regular group meetings organized by microfinance organizations,

but even in the “mandatory” treatment of Karlan and Valdivia (2011), attendance rates are only 88 percent. Screening on initial interest in training is no guarantee of high take-up rates. Bruhn and Zia (2012) and Valdivia (2011) both focus on samples who had initially expressed interest in attending a training course, but still only manage attendance rates of 39 percent and 51 percent respectively. In most short courses there has been very little drop-out conditional on attending the first session of the course, but longer courses have experienced more drop-out along the way. For example, Valdivia (2011) reports that only 42 percent of those that started the training completed at least 20 sessions, and only 28 percent at least 30 sessions.

The challenge of take-up is further illustrated through the experiences of two planned randomized experiments in which insufficient take-up prevented the experiment from occurring. Sonobe et al. (2010) planned a business training experiment among enterprises operating in a manufacturing cluster in Kenya, but had only 40 of 120 invited enterprises show up to the opening ceremony (33% take-up), preventing them from randomization based on over-subscription. McKenzie and Weber (2009) discuss a planned experiment among female SME clients of a bank in Uganda. Take-up was only 45%, and the bank ended up inviting some of the control group firms to the training to make up the planned training numbers, resulting in insufficient power for the experiment.

#### *Training content*

Table 4 attempts to summarize the key topics taught in the different courses. They all focus on general business skills that should be broadly applicable to most businesses, rather than technical knowledge or sector-specific content. However, there is significant variation in the depth and breadth of topics. The most common set of topics center around keeping business records and encouraging small business owners to separate household and business finances. There is significant overlap here with financial literacy courses, although Drexler et al. (2012) appear to be the only courses that also focus on encouraging savings and discussing debt management, which are other common financial literacy topics. Many courses, especially those targeted at potential rather than existing business owners, focus on generating a product idea and the steps needed to take it to market. A core set of topics for attempting to grow existing businesses includes marketing, pricing and costing, inventory management, customer service, and financial planning. Since few of the microenterprises have employees, employee management is not a significant part of most courses. Courses focused on larger firms have included content on quality management, lean production, and/or Kaizen and 5S techniques for continuous production improvement. Finally, in addition to attempting to business practices, some courses attempt to change entrepreneurial attitudes or aspirations. The amount of course time devoted to these aspects has been relatively light in the courses studied by economists, but Glaub and Frese (2011) review a number of non-experimental studies in developing countries in which training programs have focused much more heavily on strengthening of psychological factors.

These different types of content might be expected to impact on business performance in different ways. Simple accounting practices and financial literacy training may give business owners a better understanding of what is going on in their business, but have little short-run effect on their sales or profitability. However, in the longer term, the use of such practices may enable them to reinvest more in

their firms because of higher savings, and to make adjustments to product lines if they have a better understanding of which products are more profitable. In contrast, some of the other practices may show impacts more quickly – better marketing and customer service may directly increase sales, while costing and quality control practices may reduce costs and increase profits. Development of a new product idea could have rapid and long-lasting benefits even if no other additional practices are introduced. Changes in entrepreneurial attitudes may have impacts on how hard the owner works, and in the way they start thinking about a host of business decisions. However, because all of the available training experiments contain a mixture of different content, existing studies are unable to determine which components of training in particular matter most.

### **3. Challenges in Measuring Impact**

Impact evaluations which measure the effects of business training programs on business performance rely primarily on survey data to measure outcomes. In order to obtain credible and useful estimates, studies must have sufficient power, measure impacts over a time horizon that is sufficient for impacts to be felt, deal with survey attrition and selective survival and start-up of firms, and consider the possibility that training changes how firms report business outcomes even if it doesn't change those outcomes. We discuss each of these challenges and assess how well existing studies have met them in order to provide context for the results discussed in the next section.

#### *Power*

The power of a statistical test is the probability that it will reject a null hypothesis given that the null hypothesis is false. A starting point for most business training evaluations is to test the null hypothesis that the intervention had no effect, so the power of the experiment is a measure of the ability to detect an effect of training if such an effect does exist. The key determinants of the power of a study are the size of the sample, how much heterogeneity there is in the sample (the more diverse the set of firms the harder it is to measure a change for them), whether the intervention takes place at an individual or group level (power is lower for a given sample size when treatments are allocated at the group level), and the size of the treatment effect. Low take-up rates dilute the treatment effect, reducing power.

Table 5 compares different studies in terms of these components of power. Most studies conduct individual level randomization, with the exception of several studies working with microfinance groups, where different repayment groups are allocated to treatment or control. A typical study involves approximately 200 to 400 individuals or groups being treated and similar numbers in the control group, although sample sizes have been smaller for the studies based on specific industrial clusters (Mano et al, 2012; Sonobe et al, 2011). A useful summary statistic of the cross-sectional heterogeneity in baseline firm outcomes is the coefficient of variation (C.V.) of profits or revenues, which is the ratio of the standard deviation divided by the mean. The two studies with lowest C.V.s are both studies which restrict the heterogeneity in firms eligible for the study – de Mel et al. (2009) required firms to have baseline profits below 5000 Rs per month (US\$43), while Berge et al. (2011) restricted training to firms with loan sizes in a reasonably narrowly defined range. In contrast, most studies of microfinance clients or firms in an industrial cluster contain a much wider mix of firms, resulting in coefficients of variation

which can exceed 2 or more. The more different the firms are, the harder it will be to detect changes in their average arising from treatment.

To give a sense of the power of these different studies, we then calculate the power to detect a 25% and a 50% increase in profits and revenues, given the C.V., sample size, and training take-up rate of each study.<sup>1</sup> We assume analysis is by an ANCOVA specification (outcome regressed on treatment and the lagged outcome variable) using a single baseline and single follow-up survey. Power then also depends on how strongly correlated the baseline profits and revenues are with the follow-up. McKenzie (2011, 2012) examines this for several firm surveys and finds autocorrelations in the 0.3 to 0.5 range, reflecting the fact that firms experience a large amount of idiosyncratic variation in their profits and revenues from one period to the next (Fafchamps et al, 2012). We assume an autocorrelation of 0.5, and ignore any survey attrition, which would further weaken power. However, we also ignore any reduction in heterogeneity that comes from stratified randomization or other methods to improve power (Bruhn and McKenzie, 2009).<sup>2</sup> Nevertheless, these calculations should provide a useful approximation of the approximate power of different studies. Group-randomized studies add a further complication for power comparisons, since their power is reduced relative to individual level randomization by a factor which depends on the intra-group correlation. Our solution is to provide a range of power, where the higher number would occur if this intra-group correlation is zero (equivalent to individual randomization) and the lower number would occur if the intra-group correlation was one (equivalent to having only one observation per group). Given the heterogeneity in firms, it is likely the intra-group correlation is relatively low, making actual power closer to the higher number.

Many funding agencies consider 80% to 90% power an appropriate target (Duflo et al, 2008) and power of 80% or more is the standard in medical trials (Schultz and Grimes, 2005). Table 5 reveals that many, if not most, business training experiments fall well below these levels in terms of power to detect a 25 percent or 50 percent increase in profits or revenues. None of the studies achieve 80% power to detect a 25 percent increase in revenues, and only de Mel et al. (2012) and possibly Berge et al. (2011) exceed 80% power for a 25% increase in profits. The industrial cluster studies (Mano et al, 2012; Sonobe et al., 2011) and Bruhn and Zia (2012) are particularly low powered. However, the importance of heterogeneity and take-up is also seen in the low power for Valdivia (2011), which has the largest sample size of any individual experiment, but relatively high heterogeneity among firms and only 51% take-up (which requires 4 times the sample size to achieve a given power as would be the case with 100% take-up).

In practice some of these studies attempt to further increase power by reducing the standard deviation of revenues or profits through truncating (capping the outcome at a specified level) or trimming (dropping observations with outcomes above a certain level) the top 1 or 5 percent of observations. This

---

<sup>1</sup> Power calculations were done using the `sampsi` command in Stata, and a test size of 0.05. See McKenzie (2012) for more discussion of the validity of the underlying assumptions needed for these calculations in firm experiments..

<sup>2</sup> Power can be further increased by taking multiple rounds of follow-up surveys and pooling them together to look at an average impact over the pooled survey rounds (McKenzie, 2012). However, de Mel et al. (2012) is the only business training study to do this, so we ignore this element in our power calculations.

increases power, but if the training has large effects for a few firms and small effects for most, may eliminate the very effect the study is trying to detect. Karlan et al. (2012) argue that the prospect of a “tail” event may induce firms to experiment with new methods and techniques, such as those taught in training or by consultants. Truncation should be preferred to trimming for this reason.

We should note also that power will likely be much higher for detecting binary outcomes, such as whether or not a new business is started, whether or not a firm applies for a loan, or whether or not it implements a particular business practice. Studies with low power to inform us about the impacts of training on ultimate business outcomes may therefore still be informative about other impacts of training.

### *Timing of Effects*

The short and long-run impacts of many policies can differ substantially, so a key challenge for impact evaluation is figuring out when to measure impacts (Woolcock, 2009). For business training, one might expect firms to make some changes to their firms relatively quickly after training, but for the full impact of training to take some time. Impacts on business survival may also take time to materialize. However, firms may also start some practices and then drop them, so that surveys which measure what is taking place in the business only several years after training may miss the period of experimentation. Ideally, then studies should trace the trajectories of impacts, measuring both short and longer-term effects.

Table 6 provides details on the number of follow-up surveys, their timing, and their attrition rate for the different studies. All but one study (de Mel et al, 2012) currently only take a single follow-up survey, providing a snapshot of information on the training impact, but no details on the trajectory of impacts. Five of the eleven studies are very short-run, looking at impacts less than a year after training has finished. Another three look at impacts at exactly one year after training, while only de Mel et al. (2012) and half of the sample of Karlan and Valdivia (2011) are surveyed at two years after training. De Mel et al. (2012) find that the impacts differ in the short and medium term in their study. For example, examining impacts within the first year would show that business training for women out of the labor force led to large impacts on business entry, whereas surveys 16 and 25 months after training show the control group had caught up in terms of business ownership rates.

### *Survey Attrition and Selective Survival or Start-up*

Survey attrition is another problem complicating inference, especially if the reasons for attrition are business failure, refusal because of disappointment with the training effects, or because successful business owners have moved out of the area. Survey attrition rates can be related to the timing of the follow-up surveys – a longer interval elapsing since baseline can make it more likely people have moved or that contact details have changed, but may also help reduce the chance that the control group attrits because they were disappointed not to be selected for training. Attrition rates range from as low as 5.3 percent in Field et al. (2010) and 6 to 8 percent in de Mel et al. (2012), up to 24 percent in Karlan and Valdivia (2011) and 28 percent in Klinger and Schündeln (2011). It is perhaps no coincidence that the highest attrition rates come from studies conducted in Latin America, our personal experience has been

that in many Latin American countries firms are less willing to answer surveys than in Africa or South Asia.

In addition to attrition out of the survey, attempts to examine impacts of training on business outcomes face additional difficulties when training influences the rate of business survivorship, or the likelihood of business start-up. For example, Giné and Mansuri (2011) find that training led to a 6.1% reduction in the rate of business failure for male owners (significant at the 10% level only). If training led to relatively unsuccessful firms in the treatment group staying in business who would have closed had they been in the control group, then a straight comparison of profits or sales by treatment status would understate the impact of training. They therefore try a bounding approach to try and deal with this issue, which has been the approach used in several other studies. Even if training has no impact on the rate of business survivorship or start-up, it may affect the characteristics of who survives, again requiring authors to use non-experimental methods to attempt to deal with this selectivity. For example, de Mel et al. (2012) find that training (and grants) led to changes in the characteristics of who had opened a business, even though the rates of ownership were no different across treatment groups. They therefore use a generalized propensity score to reweight their regression estimates to correct for the selectivity they find on observables like ability and wealth.

#### *Measurement changed by Training*

A final challenge for measuring the impact of business training on business outcomes is how to measure those outcomes. Start-up and survivorship are objective measures that can at least be somewhat verified, whereas measuring business practices, profitability, and revenues for most firms is more difficult. Business practices (e.g. keeping accounts, separating business and household expenses, advertised in the past month) are normally relatively easy concepts for firms to understand, and questions that firm owners are usually willing to answer. However, Drexler et al. (2012) note the possibility that treated individuals may report doing certain behaviors (e.g. separating personal and business accounts) because the training told them this was important, rather than because they actually do this.

Measuring profits and revenues poses further problems. Many small business owners do not keep written records of these items, and owners of larger firms who do keep records may be reluctant to share them (because of concerns about theft, or because of taxation issues). De Mel et al. (2009a) study several approaches to obtaining profits from microenterprises and conclude that, in their context at least, a simple direct question is more accurate and much less noisy than attempting to calculate profits from revenues and expenses. However, collecting profits has proved difficult for many studies, with several studies not collecting profit data at all (Valdivia, 2011; Giné and Mansuri 2011; Klingler and Schündeln, 2011), collecting it but not using it due to too much noise (Drexler et al, 2012), or collecting only profit margins on the main product, rather than overall profits (Karlan and Valdivia, 2011).<sup>3</sup> Most studies have collected revenue data (Bruhn and Zia, 2011 is an exception), but some have struggled with much lower response rates for revenues than for non-financial business questions (e.g. Drexler et al,

---

<sup>3</sup> In addition profits are not examined in two studies which look at business start-up as outcomes – Field et al. (2010) and Premaud et al. (2012).

2012 have a 46 percent attrition rate on revenues compared to 13 percent for their questionnaire as a whole). Similarly, Bruhn and Zia (2011) note that about one-third of clients in their study did not provide profit data.

Even when studies are able to obtain data on profits and sales, a concern for business training experiments is that business training changes the reporting of this data, irrespective of whether or not it actually changes profits and sales. Whether or not this induces a bias depends on whether business owners systematically under- or over-report their financial data in the absence of training, otherwise training may just reduce the variance of these outcomes but not change the mean.<sup>4</sup> Few studies to date have discussed or attempted to address this issue. Exceptions are Drexler et al. (2011), who look at reporting errors (e.g. reporting profits higher than sales, or bad week sales higher than average sales) to see whether treatment reduces these reporting errors, and at the difference between self-reported profits and profits calculated as the difference between revenue and expenses; Berge et al. (2011), who compare self-reported profits to revenue minus expenses for treatment versus control groups; and de Mel et al. (2012) who do the same, and who also control for detailed measures of accounting practices as a further robustness check. De Mel et al. (2012) find little evidence that training has changed reporting, whereas Drexler et al. (2012) find that their rule-of-thumb training reduces the number of errors in reporting, and Berge et al. (2011) find that training increases the gap between self-reported profits and revenue minus expenses.

#### **4. Impacts of Business Training Interventions**

The previous section highlights the issues with statistical power, attrition, measurement, and timing of follow-up that present challenges for interpreting the impacts found in the different studies. With these caveats in mind, we examine the extent to which business training has been found to have impacts on business start-up and survivorship; business practices; business outcomes; and outcomes for microfinance lenders. Since studies of other microenterprise interventions (e.g. De Mel et al, 2009b) have often found differences by gender, we separate results by gender to the extent possible.

##### *Impacts on Start-up and Survivorship*

Table 7 summarizes the impacts of different studies on business survivorship and new business start-ups. The coefficients are marginal effects on the probability of either outcome occurring, so that a coefficient of 0.06 can be interpreted as a 6 percentage point increase. Consider first the impact on business survival. Survivorship is difficult to examine when attrition rates are high, since firms which close down may be more difficult to interview, and bounds which allow for attrition can be very wide. The only study with a significant effect at the 5% level is Mano et al. (2012), who find a 9 percentage point increase in the likelihood of survival 12 months after their training. They do not provide bounds for this effect that control for survey attrition, but note that none of the training participants had closed down. Giné and Mansuri find a 6 percent increase in the likelihood of survival 18 to 22 months post

---

<sup>4</sup> A related concern is that people who take training may overreport profits or revenues after training to exaggerate how well their firms have benefited from training. The same robustness checks as described in the text will also help rule out this sort of behavior; as well as detailed probing and observation from the surveyors.

training for the male owners in their sample which is significant at the 10% level, but no change for female owners, while Valdivia (2011) actually finds training to have a marginally significant reduction in the likelihood of survival for female firm owners, which he attributes to the training teaching owners to close down losing firms. The remaining studies which do report survivorship find insignificant impacts, but with confidence intervals wide enough to include at least a 5 percentage point increase or decrease. The sample-weighted average impact is -0.03 for female-owned enterprises and +0.04 for male-owned enterprises, although given the differences in timing of follow-up across firms, it is unclear how useful an aggregated impact is.

Studies which focus on existing firm owners sometimes look at start-up of a second business, but do not find significant impacts. The studies which focus on training specifically tailored at starting new businesses have mostly found some impacts. Klinger and Schündeln (2011) find very large point estimates for the effect one year later of taking part in the second phase of Technoserve's business plan competition in which training occurs, although the confidence intervals are very wide, and this impact also includes the joint impact of grants given to the winners. Premand et al. (2012) have a sample of 1500 youth and find taking part in an entrepreneurship track instead of academic track in the final year of university leads to a 6 percent increase in self-employment rates for males and 3 percent for females one year later.<sup>5</sup> Field et al. (2010) examine whether women report any personal labor income over the past week 4 months after training, which is a combination of an effect on business start-up and an effect on survival. They find that upper caste Hindu women who took the training were 19 percentage points more likely to report having income, whereas the training had no effect on lower caste Hindu women or on Muslim women. They attribute the lack of impact on these other groups to impacts of social restrictions, arguing that training helped women whose businesses had been held down by social restrictions, but that women who faced more extreme restrictions could not respond to training.

Training therefore appears to have some success in generating short-run impacts on business start-up. However, this does not necessarily increase employment, since it may come from people switching from wage work. Indeed Premand et al. (2012) and de Mel et al. (2012) both find that short-run increases in self-employment from training are coupled with reductions in the likelihood of wage work, so that net employment effects are insignificant. Moreover, it is unclear whether this training merely speeds up the rate of entry, or permanently increases it. De Mel et al. (2012) find that training alone increases the rate of business ownership among a group of women out of the labor force by 9 percentage points within 4 months of the training, and giving them a grant increases this effect to 20 percentage points. However, by 16 and 25 months after training, the control group had caught up. Given the short time horizon of the other studies which have found start-up impacts, it is unclear whether they do would show these effects dissipating over longer time horizons.

Relatively few studies have looked at how training affects the selectivity of who starts up a business or of which businesses survive. The available evidence suggests training may enable less analytically able and poorer individuals to start businesses, and may prop up the survivorship of less profitable

---

<sup>5</sup> This effect also includes the impact of seed money given to the top placed business plans, but the authors argue via various checks that the impact is not being driven by these grants.

businesses. De Mel et al. (2012) find that training leads to an increase in start-up among women with low Raven test scores (a measure of analytical ability), and actually reduces the rate of ownership among women with the highest test scores. Giné and Mansuri (2011) find that businesses that fail tend to be ones with lower land holdings and lower digitspan recalls, but do not examine whether training differentially affects who survives. Since many studies look over only a short time frame, rates of business failure are often low, and so this issue is more important in contexts where impacts are measured over longer time frames, or where business failure rates are high – for example, Bruhn and Zia (2012) find 36 percent of businesses shut down during their study period in Bosnia, with this high rate likely due in part to the downturn caused by the global economic crisis.

### *Impacts on Business Practices*

A first link in the causal chain from business training through to business profitability and growth is that business training has to improve the knowledge of business owners and result in these owners implementing at least some of the practices and behaviors that training has taught them. One can think of other potential mechanisms through which training may affect business outcomes (e.g. through changing attitudes or work hours), but failure to find any change in practices should cast doubt on the ability of the training to improve firm outcomes.

Table 8 summarizes the impacts found by different studies on business practices. Almost all studies find a positive effect of business training on business practices, although the effect is often not significant once the sample is reduced by splitting it by gender. Studies differ in both what practices they measure, how comprehensively they measure them, and how (if at all) they aggregate them. Several studies only measure three basic practices, such as Mano et al. (2011) who record whether or not the firm keeps records, whether it analyzes them, and whether it visits customers. Others record a broader range of practices, including different types of recording, different marketing activities, and other specific practices taught in the training.

One common approach to aggregating different practices is to follow Kling et al. (2007) and normalize each practice as a z-score (subtracting the mean and dividing by the standard deviation), and then averaging these z-scores. A coefficient of 0.03 as in Karlan and Valdivia (2011) is then interpretable as an impact of 3 percent of a standard deviation. This is useful for considering the magnitude of the increase in relative terms, but does not provide much guidance as to the absolute size of the effect. Alternatively one can look at the percentage point increase in the likelihood of particular practices being implemented, or the number of practices increased out of some total, both of which provide more guidance on the absolute magnitude of the increase.

Many studies find baseline levels of business practices which are relatively low. For example Giné and Mansuri (2011) report only 18 percent of firms at baseline record money taken from the business and only 18 percent record sales. Even among larger metalwork firms, Mano et al. (2012) report only 27% of their sample keeping business records, and 20% visiting customers at baseline. Although most studies find significant increases in the use of business practices taught during the training, the magnitude of these effects, while sometimes large in relative terms, are often small in absolute terms. For example,

Drexler et al. (2012) find that the use of rule-of-thumb training leads individuals to be more likely to report separating personal and business expenses, keep accounting records, and to calculate revenues formally, with each of these measures increasing 6 to 12 percentage points relative to the control group. In Giné and Mansuri (2011), treatment impacts include a 6.6 percentage point increase in recording sales, and 7.6 percentage point increase in recording money taken for household needs. In de Mel et al. (2012), existing enterprises implement 2 more practices out of 29. There are exceptions – for example, Mano et al. (2012) find a 30 percentage point increase in the percentage of firms keeping records in the treatment versus the control group – but in general, if the magnitude of the changes in business practices is relatively small, we might expect to find it hard to detect impacts of these changes on business outcomes.

### *Impacts on Business Profits and Sales*

Ultimately from the viewpoint of an individual firm owner, training must result in increases in profitability in order for it to be worth the investment. However, as noted previously, many studies have struggled with measuring profitability, and so not all studies consider this as an outcome. Table 9 summarizes those which have done so, converting point estimates which are sometimes in terms of profit or sales levels to percentage increases in the control group mean to enhance comparability across studies. Several studies examine gender heterogeneity by reporting a point estimate for males, and then an interaction effect for females, but no test of the overall impact on females which is why confidence intervals are sometimes shown for males and not females in this table. Often studies have more than one specification for profits and/or revenues, with variation in whether or not they include different controls, and whether or not they truncate or trim the data, or take a log transform. We report impacts on the measure which corresponds most closely to profits or sales in the last month.

We see that few studies have detected significant impacts of business training on business profits or sales, although the confidence intervals are very wide in many cases. The studies that have most power according to the calculations in Table 5 are the ones most likely to show significant effects. Berge et al. (2011) find training to increase profits for males by 24 percent and sales by 29 percent, with impacts which are insignificant and close to zero for women. Their study has 18 percent attrition, and so to examine the sensitivity of these estimates to attrition they use bounds. The lower bound for the male profit impact is an 8.7 percent increase, which is not statistically significant, which coupled with the short time horizon of the follow-up survey (5 to 7 months), suggests some caution in using this estimate to conclude that training has had large increases in profits. De Mel et al. (2012) is the other study with most power. They find no impact of training alone on profits of existing firms over either the short or medium run, but do find significant impacts of the combination of training and a grant on short-run profits, with these gains dissipating over time. Training is found to significantly increase profits and sales of the businesses started by women who were out of the labor force by approximately 40 percent, although the confidence intervals are wide around this level. Valdivia (2011) is the only other study to find significant impacts on revenues, finding a 19 percent increase for the treatment group that got both training and intensive one-on-one technical assistance. One potential concern in interpreting this estimate is that the coefficient changes dramatically when covariates are added to the regression, suggesting a lack of balance in the follow-up sample, perhaps due to selective attrition.

Several studies have emphasized the possibility that business training may have its strongest impacts on sales during a bad month. The working paper version of Karlan and Valdivia (2011) stressed this avenue, noting that training might help clients identify strategies to reduce downward fluctuations in sales by thinking about diversifying the products they offer as well as being more proactive about alternative activities during slow months. Their difference-in-difference estimate finds a 30 percent increase in sales during a bad month. However, the published version of the paper de-emphasizes this impact, noting that when their preferred ANCOVA specification is used instead of difference-in-differences, the impact falls to an insignificant 5 to 7 percent increase. This possibility of impacts during bad times is also emphasized by Drexler et al. (2012), who find that their rule of thumb training leads to an increase in sales during bad weeks which is significant at the 10 percent level. However, Drexler et al. also ask firm owners to report sales in a bad month, instead of a bad week, and find a very small and insignificant impact of training on this measure. Giné and Mansuri (2012), de Mel et al. (2012), and Valdivia (2011) also find no significant impacts of training on sales during bad months. Viewing these studies together therefore leads us to conclude that the evidence that training has particularly strong effects during bad periods is rather weak.

Since the costs of training often increase less than proportionally with the size of the firms taking training, subsistence microenterprises would likely need to see much higher percentage increases in profits from training to make training worthwhile than would larger firms. For example, a firm earning \$1 per day would need to see a 13.7 percent increase in profits for the cost of \$100 of training to be recouped over two years, whereas a firm with \$500 in monthly profits would only need a 2 percent increase in monthly profits to recoup the costs of training which costs \$250 over two years. We see that the confidence intervals for the studies which do look at profits are all wide enough to include these rates of return. If we add an opportunity cost of funds and risk aversion, the desired returns would be higher, but the basic point remains that, for larger firms in particular, a percentage increase in profits that is too small for existing studies to detect as different from zero may be enough to justify the costs of training. This relies on the increase in profits lasting for at least a couple of years, which requires measurement of the trajectory of impacts.

#### *Impacts on Microfinance institution outcomes*

Since many of the studies have worked with microfinance clients, they have also looked at outcomes using administrative data from the microfinance organization. This data has the advantage of potentially being available with less attrition and over longer time periods, and is useful for assessing whether offering training is cost-effective for the microfinance organization, but is less useful for telling us about how such training impacts on firms. Karlan and Zinman (2011) find training resulted in a 4 percentage point increase in client retention rates, and 2 to 3 percentage point increase in the likelihood of perfect repayment (although this is only marginally significant). However, they also note that some of the clients who do leave cite the added length of the weekly meetings due to the training sessions as a factor in dropping out of the program. They note that these benefits appear to make the training profitable from the lender side, and, after their study, FINCA implemented the mandatory version of their training in all village banks.

Giné and Mansuri (2011) find training to lead to a 16 percent increase in loan size for males, a reduction in loan size for females, no change in repayment rates, and a change in the selection of who borrows, with individuals with higher predicted probabilities of default being less likely to borrow after training. Field et al. (2010) find upper caste Hindu women who take training are 13 percentage points more likely to borrow. In contrast, Drexler et al. (2012) and Bruhn and Zia (2012) find no significant impacts of training on the likelihood of taking loans or loan size, although Bruhn and Zia find an increase in loan duration and refinancing of loans, which they attribute to trained individuals making longer-term investments and being more aware of interest rates available.

## **5. Non-Experimental Studies**

The studies shown in Tables 1 and 2 are all randomized control trials. There are also a series of studies using non-experimental data and attempting to correct for selection in other ways. We would put somewhat less weight on even the best of the non-experimental studies because of concern for selection bias. However, given the scarcity of training impact studies, and the multi-dimensional heterogeneity discussed above, we review the non-experimental studies with more plausible identification. Moreover, given the costs of running RCTs relative to the cost of non-experimental studies, funders often ask whether a combination of RCTs and non-experimental studies might be a more cost-effective way of learning, especially with respect to learning about the heterogeneity of impacts. We thus spend some time discussing three representative non-experimental studies. These studies demonstrate both the potential and the challenges of the non-experimental approach. We then discuss the cost advantages of these studies relative to RCTs.

Mckernan (2002) conducts “the first measure of the non-credit effects of microcredit programs” (p. 93) by examining data from 1800 households in 87 villages in Bangladesh. The MFIs operating in these villages (Grameen, BRAC, and BRDB) offer credit bundled with “non-credit services.” The non-credit services are sometimes provided directly by the lenders – e.g., vocational training or health and other educational programs – but also come from the fact that group lending induces monitoring by other microenterprise owners. The analysis is based on observational data rather than an RCT. Mckernan’s premise is that credit operates through capital stock, and the bundle of non-credit services affects output or income conditional on a given level of capital. So, adding a control for capital stock to a regression of profits on access to microfinance comparing regressions controls for the ‘finance channel’ of microcredit. The idea, then, is that any remaining effect of microcredit must be due to the bundled services. Mckernan faces several challenges in the analysis. Most seriously, both the level of capital stock and participation in microlending are endogenous. The paper attempts to deal with these using both variation in the presence of MFIs at the village level, information on land ownership in the extended family of the household heads, and an eligibility rule of Grameen Bank which states that borrowers must own no more than 0.5 hectares of land. Using these instruments, she finds that the total effect of the program is to increase self-employment profits by 175 percent. The majority of this increase – 125 percent of self employment profits – is assigned to the non-credit aspects of borrowing. As Mckernan notes, there are potential issues with all of these instruments. But even if we accept the instruments as valid, we face a second serious issue from the perspective of learning about training and capacity building: the “non-credit interventions” are a bundle of many interventions. These include

training but also, for example, incentives for effort that might be generated by group lending. Because lending group members are affected by whether other members repay their loans, each group member has an incentive to encourage effort by other group members. This additional effort may lead to an increase in profits at any given level of capital. Thus, how much of the measured 125 percent increase in profits is due to training or capacity building alone is impossible to determine.

Biosca et al (2011) also examine the effect of bundling microfinance with business development services and health information programs in the state of Chiapas, Mexico. The business training offered in the program they study was based on the Freedom From Hunger program and administered in 30-45 minute blocks at regular biweekly meetings. The authors employ a “pipeline” strategy first used in the microfinance literature by Coleman (1999). As with McKernan, the challenge is to identify a valid comparison sample against which to measure the program effects. In the pipeline approach, current recipients of the program (the “treatment group”) are compared to individuals in communities which will be eligible for the program in another year’s time. In this case, the MFI identified (in ways that are not detailed in the paper) borrower groups who would participate in the bundled services when they were offered in the ‘control’ community at a later point in time. Though there is clearly effort to identify a valid comparison group here, there are several challenges with the approach. First, the rollout across villages is usually not random. Second, even if the training is made compulsory by the lender, they may be dropout by borrowers after the first lending cycle. Identifying who among the comparison group would drop out is often difficult. In the case of the BDS program studied by Biosca et al, for example, borrowers in two of the MFI’s branches were asked if they would participate in the training programs if offered the chance. On the one hand, the percentage responding affirmatively to the hypothetical question (71 percent) was comparable to the percentage actually participating in the ‘treatment’ branches (65 percent). On the other hand, we know from other studies that individuals who say they would participate and those who actually do participate often differ. Business development services are found to lead to a 25-30 percent increase in household income – though the effect is not statistically significant once municipal fixed effects are included – and to make it less likely that a household is below the poverty line measured by assets.

Bali Swain and Varghese (2011) assess the impact of training provided to self-help groups (SHGs) in India. The training is sometimes provided by government workers and sometimes by NGOs. They must deal with selection at two levels: the decision to join an SHG, and the decision to take training conditional on joining an SHG. The authors control for the selection into the SHGs by comparing members of SHGs with a group of individuals who have signed up for an SHG but have not yet received a loan. (See Coleman 1999 for a comparable strategy.) A well-known problem with this approach is that individuals may borrow for only one cycle and then drop out from the group. There will be no similar dropout from the ‘control’ group. Bali Swain and Varghese note that dropout rates on average are just over 8 percent, which suggests this is an issue, but perhaps not a mortal issue. Training is administered through the SHGs, but both new (pre-credit) and established members of SHGs receive training. Hence, if proper comparison groups can be identified, the authors can measure the effects of training in the presence and absence of credit. However, the training is offered to SHG members in a non-random

manner, leading to a second selection issue. The authors use propensity score matching methods to identify a comparison group.

## **6. Boosting the intensity and working with larger firms**

As noted in Sections 2 and 4, many of the training sessions are relatively short in duration, and the increase in business practices has been relatively small in a number of studies. One response to this is that what is needed is more in-depth and individualized follow-ups on the training, while another has been to also focus on larger firms where management practices may potentially be of greater importance. We discuss the results of studies which have pursued these two approaches.

### *Individualized follow-ups*

Three of the business training impact evaluations have also had a treatment group which followed up on the classroom training with individualized follow-ups. Drexler et al. (2012) had trainers visit 8 times over five months to answer any queries, and verify and encourage the use of accounting books, and correct any mistakes in completing books. They find no significant effects of this additional follow-up. Giné and Mansuri (2012) have “hand-holding sessions” in half of the community organizations, where firms get four months of being visited 1-2 times per month, to discuss topics learned, answer questions, and suggest solutions to potential problems. They also find this hand-holding had no effect on any of their aggregate outcomes for either men or women.

In both these cases, the follow-ups seem mostly intended to reinforce the general business skills taught in training, rather than to provide firm-specific individualized advice. Valdivia (2011) also examines more intensive follow-up, but in his case, has trainers also providing more specific technical assistance tailored to the needs of the women’s businesses. It combined individual sessions and visits with group sessions among small groups of similar businesses over a three month period, although the paper does not provide details on how many hours of follow-up firms received. He finds some evidence to suggest this technical assistance helped firms – women assigned to receive this technical assistance experienced a 19 percent increase in revenue (significant at the 10 percent level) and showed more improvement in business practices than women who were assigned just to the basic training. The cost of this additional effort is not provided in the paper, so it is unclear what the cost-benefit of this more intensive follow-up is.

### *Individual consulting*

A related literature looks at the impact of providing consulting services on a one-on-one basis to firms in order to improve business and management practices. The closest to the business training experiments is the work of Karlan et al. (2012), who study a mix of 160 male and female tailors in Ghana with 5 or fewer workers. They used local consultants from Ernst and Young in Ghana, who met with the tailors in visits of 30 minutes to 1 hour several times a month over one year, with the average firm receiving 10

hours of consultant time over a year, at zero cost to the firm. In addition some firms receive grants of around US\$133. They find some of the consultants recommendations were adapted for some months, but had been abandoned one year after training stopped. There is no significant impact of either treatment on profits or revenues, with some specifications showing negative effects in the short-run, although power is very low and confidence intervals wide.

Bruhn et al. (2012) evaluate a state government program in Puebla, Mexico, which paired small businesses with a consultant from one of several local consulting firms, and had these consultants spend approximately 4 hours per week over a year assisting the firm in overcoming constraints to growth. 432 firms applied to the program, and 150 were chosen to receive heavily subsidized consulting services (the cost was about 10% of the commercial rate). The mean number of employees was 14, and 72 percent were male-owned. A single follow-up survey one to three months after the consulting was used to assess impacts on the firms, with the authors finding large point estimates for the impacts on sales and profits, which are sometimes significant depending on the measure used and the extent of trimming. However, the study faces many of the same challenges as the business training studies reviewed above. First, the firms applying are very heterogeneous, with the baseline coefficient of variation of sales being 3.7, and still being 2.4 after trimming the top 1%. Second, despite all firms signing a statement of interest, only 80 out of the 150 firms assigned to treatment decided to take it up (53%). Third, attrition rates are reasonably high and there is additional item non-response on profits and sales even among those who are interviewed, so that only 288 firms (66.7%) provided data on profits in the follow-up survey. The combination of high heterogeneity and reduced take-up lowers power, which combined with the high attrition rates would mean any bounding exercise would give non-informative bounds for impacts on profits and sales. These are likely to be challenges facing any similar government program offering subsidized consulting or business services to firms, such as the matching grant programs used in many World Bank private sector loans.

The final individualized consulting study is Bloom et al. (2012), who focus on a much smaller sample of 17 large textile firms in India, with the typical firm having 270 employees, 2 plants, and sales of US\$7.5 million per year. They provide 11 of these firms with five months of free intensive consulting provided by Accenture Consulting, with the consultants averaging 781 hours per treated plant on average, spending their time attempting to get the firms to implement 38 key management practices related to quality control, factory operations, inventory, human resource management, and sales and order management. They deal with the problem of small sample size by focusing on very homogeneous firms and collecting lots of data from them, including weekly data on quality, output, and inventories. They find that adopting these management practices raised productivity by 17% in the first year through improved quality and efficiency and reduced inventory, and some evidence that within three years this led to the opening of more production plants. The results thus show that in large firms at least, intensively changing management practices can lead to substantial improvements in firm performance. However, they can only indirectly estimate the changes in profits from this effort.

## **7. What we don't know**

There are now a range of studies which look a variety of business training programs, and examine impacts on business practices, business outcomes, and potentially outcomes for microfinance institutions. However, there are still a number of open questions that existing studies have barely touched upon, but that are important for thinking about the case for policy action to support business training.

#### *Who does training help most?*

Our discussion above has touched on heterogeneity in outcomes by gender of the owner and, to some extent across studies, by firm size. Several studies have examined heterogeneity in other dimensions such as owner's education and baseline business skill levels, business sector, and interest in training. However, the low power of most studies to find average effects for the full sample also means there is relatively little power for examining heterogeneity in effects.

As a result, it is still an open question who benefits most from training (or which types of training are most suitable for which firm types?). On one hand, poor subsistence firms whose owners are only running the business because they cannot find a wage job may have very low business skills suggesting that it should be relatively easy to make improvements, but perhaps the owners will be less interested or able to implement the practices taught, or many of these practices may only start making a business as businesses reach a larger scale. There is much talk of targeting gazelles- firms that grow rapidly – but even if characteristics to identify such firms in advance can be found, it is unclear whether these firms need the help, or would grow fast anyway. Theoretically one would like to target firms where skills are the binding constraint on growth, but there is little evidence to date to say who such firms are, especially among the more micro firms.

#### *How does it help them – and do any gains just come at the expense of other firms?*

Most studies have done relatively little beyond looking at business practices to examine the channels through which training has impacts, in part because of their low power in detecting an impact on profits in the first place. For example, does training enable firm owners to use the same inputs more efficiently (thereby reducing costs and wastage), or is the main impact through increasing marketing and sales efforts thereby increasing revenues at the same cost ratios? The impacts for policy will differ depending on the channels. In particular, one possibility is that gains to the treated firms come as a result of them taking customers from other firms. For example, Mano et al. (2012) note that their training encouraged participants to be proactive in getting new customers, and participants began issuing invoices and receipts with their mobile phone numbers printed on them, which they believed had helped boost sales. Berge et al. (2011) suggest their training had impacts on male clients through an expansion effect, since profit margins did not change whereas sales did.

Such spillovers have possible implications for both internal and external validity. If the increased sales come mainly as a result of taking business from the control group firms, then this a violation of the stable unit treatment value assumption (SUTVA), and the experimental estimate will no longer give the average impact of training for the sample population. If the increased sales come mainly from other firms not in the sample, the results of the experiment could be misleading of the gain to society from

scaling the training program up. It should be noted that spillovers might also be positive, if control or non-sample firms copy some of the techniques or new products introduced by firms which have gone through training. Indeed this possibility is often given as one of the main justifications for public subsidies of matching grant programs which subsidize the purchase of business development services by small and medium enterprises. These issues are part of the broader question of how competition responds to newly trained firms – we do not know whether this deters some new firms from entering the industry, causes others to exit, or causes the incumbents who remain in business to make other changes to the way they run their businesses.

To investigate this issue, a much larger sample is needed, and experimental variation in the intensity of the treatment within different geographical areas could then be used to test for and measure these spillovers.

#### *Do larger impacts emerge over time?*

As noted above, most of the studies take a single snapshot of the impact of the training at a relatively short interval after training has ended. Two studies which have traced trajectories of impacts suggest that effects can indeed vary a lot of time. In de Mel et al. (2012) the impacts on business start-up fade with time, as control firms catch up. Bloom et al. (2012) find that introducing management practices in larger firms shows immediate effects on quality, then slowly leads to changes in inventory levels, output and productivity, and it is only after several years of using these practices that impacts start to show in terms of employment generation (through new plants being opened). Given the large interest of many governments in employment creation, studies which look only within a year or so of treatment ending may miss effects that take some time to be realized – or conversely, we may find that effects which look very promising in the short-term dissipate over time.

#### *If it is so helpful, why don't firms purchase it?*

It is noticeable that all the business training studies reviewed here offer the training for free, as do two out of the three consulting experiments with the other offering a 90 percent subsidy. In part this is done for research purposes – in order to ensure sufficiently high take-up and to provide evidence on how training influences a range of firms. Even then we have seen limits to demand, with some studies struggling to get people to undertake training even when it is free.

As a result, we know very little about what types of firms would choose to purchase training at market prices, and the effects of training on this subgroup of firms. Moreover, we know very little about whether there are market failures that prevent firms who would benefit substantially from training from purchasing this training at market prices. Several potential constraints or market failures are discussed in the literature (e.g. Karlan and Valdivia, 2011; Bloom et al, 2012, Karlan et al, 2012). The first is that of an information failure, whereby the value of business training services is not understood by entrepreneurs, and may be particularly understated by those with the most to gain who do not realize how poorly run their firms are. A second is credit constraints, whereby firms may find it harder to borrow to finance training, an intangible asset, than to finance assets which can be seized by a bank in the event of non-repayment. A third is insurance market failures, whereby firm owners may be reluctant

to take training even if they think it has a high expected payoff, because they are unable to insure against the possibility that it doesn't work. A fourth possibility is supply-side constraints, whereby consulting or training services simply do not exist in the market, so that even if a firm wanted to purchase these services it would find it hard to. A final possibility is positive spillovers, whereby the public gains to training greatly exceed the private gains, causing firms to underinvest. Each of these appears plausible in at least some settings, but to date there is no evidence as to which constraints or market failures appear to occur most frequently. The optimal policy response will differ depending on which constraint binds, so making progress on the issue of why more firms don't purchase training or consulting is likely to have useful implications for policy efforts.

### *Practices or Personality?*

Much of the tradition in business training courses has been focused on teaching particular practices that firm owners can implement in their firms. However, another school of thought is that equally, if not more, important is the attitudes and personalities that business owners have. Premand et al. (2012) report that one of the main objectives of trainers in their study was to change students' personalities to "turn them into entrepreneurs". They find that their intervention did lead to measurable and significant changes in several domains of personality. For example, assignment to the entrepreneurship track led to a significant decrease in agreeableness, ranging from 0.23 to 0.42 standard deviations compared to the level of agreeableness in the control group. They cite literature suggesting that agreeableness is negatively associated with occupations such as business professionals or managers and that, as a result, note that the behavioral skills for entry into self-employment could differ from behavioral skills needed to hold a salaried job. In addition, they find some impact of training on extraversion, which they view as consistent with the emphasis in training on making students less shy and more outspoken to turn them into entrepreneurs.

There is also a range of different training courses studied by psychologists which focus more on the personality aspect of being an entrepreneur than on the specific skills (Glaub and Frese, 2011). While several studies have incorporated some aspect of aspirations or entrepreneurial attitudes into their content, to date there is no research which tests the relative contribution of both types of training.

## **8. Conclusions and Suggestions for Future Work**

The last few years have seen rapid growth in the number of randomized trials evaluating business training programs, providing a growing body of evidence in an area with large policy interest. However, to date a number of challenges have hampered how much we can learn from these studies, with methodological concerns and heterogeneity in both training content, and in the characteristics of who is trained, making comparisons across studies difficult. Many of the key questions needed to justify large-scale policy interventions in this area remain unanswered.

In order to learn more from the next generation of studies, we believe the following elements are needed.

1. Much larger samples or more homogeneous firms: rather than more studies with 100 to 500 individuals in each treatment or control group, we ideally need studies to move to samples of several thousand or more. This would increase the power of studies, and allow more consideration of what types of people training works for best. An alternative to large cross-sectional samples is to reduce the heterogeneity of the sample by focusing on firms within one industry and size category, and collecting much more frequent time series data on these firms (McKenzie, 2011, 2012).
2. Better measurement of outcomes: measuring firm profits and revenues has proved a real challenge for many studies, and even less evidence is available on how exactly training changes the firm's production process. Further efforts to improve the measurement of financial information (and to ensure that there is not just a measurement effect of training) are needed, while focus on a specific industry or sector may allow more detailed production level monitoring of physical outputs and inputs.
3. Designing experiments to measure spillovers: this could include more use of GPS measurement to measure local spillovers (Gibson and McKenzie, 2007), and randomizing the intensity of training at the local market level to see whether the effects differ when all firms competing in a local area are trained versus when only some of them are trained.
4. Measuring trajectories of outcomes over longer periods of time: the impacts of training may differ in the short and medium terms, so measuring outcomes at multiple points in time will enable better study of whether impacts take time to materialize, or whether others which emerge quickly persist.
5. Testing which elements of content matter: with larger samples, one can build on the work of Drexler et al. (2012) and test different forms of training, in order to determine which elements of business skills have most impact, and whether training needs to focus on entrepreneurial personality as well as processes.
6. Understanding market failures and building market-based solutions: almost every study has given training away for free and still experienced difficulties getting take-up. There are many open interesting questions concerning how one helps develop a market for these business services and the types of policies that can overcome market failures that prevent firms from using these markets.

## References

Bali Swain, Ranjula and Adel Varghese, 2011, "Delivery Mechanisms and Impact of Microfinance Training in Indian Self-Help Groups," *Journal of International Development*, 2011.

Berge, Lars Ivar Oppedal , Kjetil Bjorvatn, Kartika Sari Juniwaty, and Bertil Tungodden (2012) "Business training in Tanzania: From research driven experiment to local implementation", *Journal of African Economics*, forthcoming.

Berge, Lars Ivar Oppedal , Kjetil Bjorvatn, Bertil Tungodden (2011) "Human and financial capital for microenterprise development: Evidence from a field and lab experiment", *NHH Discussion Paper Sam 1, 2011*.

Biosca, Olga, Pamala Lenton, and Paul Mosley, 2011, "Microfinance non-Financial Services: A Key for Poverty Alleviation? Lessons from Mexico," Sheffield Economic Research Paper Series, No. 2011012.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (2012) "Does management matter? Evidence from India", Mimeo. Stanford and World Bank.

Bloom, Nicholas and John van Reenen (2010) "Why do management practices differ across firms and countries?", *Journal of Economic Perspectives* 24(1): 203-24.

Bjorvatn, Kjetil and Bertil Tungodden (2010) "Teaching business in Tanzania: Evaluating participating and performance", *Journal of the European Economic Association* 8(2-3): 561-70.

Bruhn, Miriam, Dean Karlan and Antoinette Schoar (2012) "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico", Yale Economics Department Working Paper no. 100.

Bruhn, Miriam, Dean Karlan and Antoinette Schoar (2010) "What capital is missing in developing countries?", *American Economic Review Papers and Proceedings* 100(2): 629–633.

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

Bruhn, Miriam and Bilal Zia (2012) "Stimulating Managerial Capital in Emerging Markets: The Impact of Business and Financial Literacy for Young Entrepreneurs", Mimeo. World Bank

Calderon, Gabriela, Jesse M. Cunha and Giacomo De Giorgi, 2012, "Business Literacy and Development: Evidence from a Randomized Trial in Rural Mexico," working paper, Stanford University.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2012) "Business Training and Female Enterprise Start-up, Growth, and Dynamics: Experimental evidence from Sri Lanka", Mimeo. World Bank.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2009a) "Measuring microenterprise profits: Must we ask how the sausage is made?:", *Journal of Development Economics*, 88(1): 19-31.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2009b) "Are Women more Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns." *American Economic Journal: Applied Economics* 1(3): 1-32.

Drexler, Alejandro, Greg Fischer, and Antoinette Schoar (2012) "Keeping it Simple: Financial Literacy and Rules of Thumb", Mimeo. LSE

- Duflo, Esther, Rachel Glennerster, and Michael Kremer. (2008). "Using randomization in development economics research: A toolkit." In *Handbook of Development Economics*, Vol. 4, ed. T. Paul Schultz and John Strauss, 3895–3962. Amsterdam, NH: North Holland.
- Fafchamps, Marcel, David McKenzie, Simon Quinn and Christopher Woodruff (2012) "Using PDA consistency checks to increase the precision of profits and sales measurement in ", *Journal of Development Economics*, 98(1): 51-57.
- Field, Erica, Seema Jayachandran and Rohini Pande (2010) "Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India", *American Economic Review Papers and Proceedings* 100(2): 125-29.
- Gibson, John and David McKenzie (2007) "Using global positioning systems in household surveys for better economics and better policy", *World Bank Research Observer* 22(2): 217-41
- Giné, Xavier and Ghazala Mansuri (2011) "Money or Ideas? A field experiment on constraints to entrepreneurship in rural Pakistan", Mimeo. World Bank.
- Glaub, Matthias and Michel Frese (2011) "A critical review of the effects of entrepreneurship training in developing countries", *Enterprise development and microfinance* 22(4): 335-53.
- Karlan, Dean, Ryan Knight and Christopher Udry (2012) "Hoping to Win, Expected to Lose: Theory and Lessons on Microenterprise Development", Mimeo. Yale.
- Karlan, Dean and Martin Valdivia (2011) "Teaching entrepreneurship: Impact of business training on microfinance clients and institutions", *Review of Economics and Statistics* 93(2): 510-27.
- Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz (2007) "Experimental Analysis of Neighborhood Effects," *Econometrica* 75: 83–119.
- Klinger, Bailey and Matthias Schündeln (2011) "Can entrepreneurial activity be taught? Quasi-experimental evidence from Central America", *World Development* 39(9): 1592-1610.
- Mano, Yukichi, AlHassan Iddrisu, Yutaka Yoshino and Tetsushi Sonobe (2012) "How Can Micro and Small Enterprises in Sub-Saharan Africa Become More Productive? The Impacts of Experimental Basic Managerial Training", *World Development*, 40(3): 458-68.
- McKenzie, David (2012) "Beyond baseline and follow-up: the case for more T in experiments", *Journal of Development Economics*, forthcoming.
- McKenzie, David (2011) "How can we learn whether firm policies are working in Africa? Challenges (and solutions?) for experiments and structural models", *Journal of African Economics*, 20(4): 600-25.
- McKenzie, David (2010) "Impact Assessments in Finance and Private Sector Development: What have we learned and what should we learn?", *World Bank Research Observer*, 25(2): 209-33.

McKenzie, David and Michaela Weber (2009) "The results of a pilot financial literacy and business planning training program for women in Uganda", *Finance & PSD Impact note no. 8* <http://siteresources.worldbank.org/INTFR/Resources/Note8.pdf>

McKernan, Signe-Mary (2002) "The Impact of Microcredit Programs on Self-Employment Profits: Do Noncredit Program Aspects Matter?," *Review of Economics and Statistics*, Vol 84(1), pp. 93-115.

Premand, Patrick, Stefanie Brodmann, Rita Almeida, Rebekka Grun and Mahdi Barouni (2012) "Entrepreneurship training and self-employment among university graduates: Evidence from a randomized trial in Tunisia", Mimeo. World Bank

Schulz, Kenneth and David Grimes (2005) "Sample size calculations in randomised trials: mandatory and mystical", *Lancet* 365: 1348-53.

Sonobe, Tetsushi , Yukichi Mano, John Akoten, and Keijiro Otsuka (2010) "Assessing the Impacts of Management Skill Training in Ghana and Kenya", paper presented at the CSAE conference, 2010.

Sonobe, Tetsushi , Aya Suzuki, and Keijiro Otsuka (2011) "Kaizen for Managerial Skills Improvement in Small and Medium Enterprises: An Impact Evaluation Study", The World Bank <http://siteresources.worldbank.org/DEC/Resources/FinalVolumeIV.pdf>

Valdivia, Martin (2011) "Training or technical assistance? A field experiment to learn what works to increase managerial capital for female microentrepreneurs", Mimeo. GRADE.

Woolcock, Michael (2009) "Toward a plurality of methods in project evaluation: a contextualised approach to understanding impact trajectories and efficacy", *Journal of Development Effectiveness*,1(1),1-14.

**Table 1: Who are the participants in business training evaluations?**

<b>Study</b>	<b>Country</b>	<b>Existing Businesses?</b>	<b>All microfinance/ bank clients?</b>	<b>Rural or Urban</b>	<b>Business Sector</b>	<b>Selected on interest in training?</b>	<b>Mean Age</b>	<b>% Female</b>
Berge et al. (2011)	Tanzania	Existing	Yes	Urban	Many	No	38	65
Bruhn and Zia (2012)	Bosnia-Herzegovina	67% existing	Yes	Urban	Many	Yes	28	35
De Mel et al. (2012)	Sri Lanka	50% existing	No	Urban	Many	No	34-36	100
Drexler et al. (2012)	Dominican Republic	Existing (a)	Yes	Urban	Many	No	40	90
Field et al. (2010)	India	24% existing	Yes	Urban	Many	No	32.4	100
Giné and Mansuri (2011)	Pakistan	61% existing	Yes	Rural	Many	No	37.6	49
Karlan and Valdivia (2011)	Peru	Existing	Yes	Both	Many	No	n.r.	96
Klinger and Schündel (2011)	El Salvador, Guatemala, Nicaragua	39% existing	No	n.r.	Many	Yes	36	28
Mano et al. (2012)	Ghana	Existing	No	Urban	Metalwork	No	45	0
Premand et al. (2012)	Tunisia	No	No	Urban	Many	Yes	23	67
Sonobe et al. (2011)	Tanzania	Existing	No	Urban	Garments	No	45	85
	Ethiopia	Existing	No	Urban	Metalwork	No	44	4
	Vietnam	Existing	No	Urban	Rolled Steel	No	40	55
	Vietnam	Existing	No	Urban	Knitwear	No	41	66
Valdivia (2011)	Peru	Existing	No	Urban	Many	Yes	43	100

*Notes:*

n.r. denotes not reported

(a) 78% of sample is existing businesses, and study does not look at business outcomes for those who were not existing at baseline.

**Table 2: Heterogeneity in Baseline Size of Firms Participating in Business Training Experiments**

Study	% with zero employees	Mean	Monthly Profits (USD)		Monthly Revenues (USD)	
		Employees	Mean	S.D.	Mean	S.D.
Berge et al. (2011)	n.r.	1.08	480	384	2102	3083
Males	n.r.	1.18	528	432	2586	2876
Females	n.r.	1.03	455	354	1847	3160
Bruhn and Zia (2012)	n.r.	2.08	700	n.r.	n.r.	n.r.
De Mel et al. (2012)	95	0.06	35	17	109	99
Drexler et al. (2012)	60	n.r.	n.r.	n.r.	747	1215
Giné and Mansuri (2011)	90	2.43	n.r.	n.r.	n.r.	n.r.
Males	86	2.51	n.r.	n.r.	380	n.r.
Females	95	2.34	n.r.	n.r.	80	n.r.
Karlan and Valdivia (2011)	n.r.	0.22	-165	4118	534	1230
Klinger and Schündeln (2012)	n.r.	8	n.r.	n.r.	6916	17333
Mano et al. (2012)	n.r.	n.r.	2200	2700	4717	5658
Sonobe et al. (2011)						
Tanzania	n.r.	5	530	1056	866	1393
Ethiopia	n.r.	33	19599	38048	142311	354163
Vietnam - Steel	n.r.	17	2627	4181	105787	98526
Vietnam - Knitwear	n.r.	20	-888	7234	7055	16509
Valdivia (2011)	n.r.	2.9	n.r.	n.r.	740	1696

Notes:

n.r. denotes not reported

**Table 3: Key Characteristics of Training Delivery**

<b>Study</b>	<b>Training Provider</b>	<b>Training content new or established?</b>	<b>Course Length (hours)</b>	<b>Participant Cost (USD)</b>	<b>Actual Cost (USD)</b>	<b>Attendance Rate</b>
Berge et al. (2011)	Training professionals	New	15.75	0	\$70	83%
Bruhn and Zia (2012)	Training organization	New	6	0	\$245	39%
De Mel et al. (2012)	Training organization	Established (ILO)	49-63	0	\$126-140	70-71%
Drexler et al. (2012)						
"Standard"	Local instructors	New	18	0 or \$6	\$21	50%
"Rule-of-thumb"	Local instructors	New	15	0 or \$6	\$21	48%
Field et al. (2010)	n.r.	New (a)	2 days	0	n.r.	71%
Giné and Mansuri (2011)	Microfinance credit officers	New (b)	46	0	n.r.	50%
Karlan and Valdivia (2011)	Microfinance credit officers	Established (FFH)	8.5-22 (c)	0	n.r.	76-88%
Klinger and Schündeln (2011)	Training professionals	Established (Empretec)	7 days	0	n.r.	n.r.
Mano et al. (2012)	Local instructors	New (d)	37.5	0	\$740	87%
Premand et al. (2012)	Govt. office staff	New	20 days +	0	n.r.	59-67%
Sonobe et al. (2011)						
Tanzania	Training professionals	New (d)	20 days	0	>\$400	92%
Ethiopia	Training professionals	New (d)	15 days	0		75%
Vietnam - Steel	Training professionals	New (d)	n.r.	0		39%
Vietnam - Knitwear	Training professionals	New (d)	n.r.	0		59%
Valdivia (2011)	Training professionals	New	108 (e)	0	n.r.	51%

**Notes:**

(a) Shortened version of existing program + new content on aspirations added.

(b) Adapted from ILO's Know About Business modules

(c) Training sessions were each 30 minutes to 1 hours, and up to 22 sessions occurred, but only half had done 17 sessions over 24 months.

(d) Based in part on ILO content + Japanese Kaizen content.

(e) Although only 42% of those attending completed at least 20/36 sessions, and only 28% attended 30 sessions or more.

FFH denotes Freedom from Hunger; ILO denotes the International Labor Organization.



**Table 4: Training Content**

<b>Study</b>	Separating household and business finances	Accounting	Financial Planning	Product ideas	Marketing	Pricing and Costing	Inventory Management	Customer Service	Business Investment & Growth Strategies	Employee Management	Savings	Debt	Using Banks	Quality	Kaizen/5S/Lean	Aspirations/Self-esteem
Berge et al. (2011)		X	X	X	X	X		X		X			X			
Bruhn and Zia (2012)	X	X	X		X				X							
De Mel et al. (2012)	X	X	X	X	X	X	X	X								
Drexler et al. (2012)																
"Standard"	X	X	X				X				X	X				
"Rule-of-thumb"	X	X									X	X				
Field et al. (2010)	X	X	X													X
Giné and Mansuri (2011)	X	X		X		X										
Karlan and Valdivia (2011)	X	X	X	X	X	X										
Klinger and Schündeln (2011)		X	X		X	X	X	X							X	X
Mano et al. (2012)	X	X												X	X	
Premand et al. (2012)		X		X	X	X										X
Sonobe et al. (2011)																
Tanzania		X			X	X								X	X	
Valdivia (2011)		X		X	X	X		X						X		X

Notes: based on training descriptions provided in research studies

**Table 5: Power of Studies to Detect Increases in Profits or Sales**

Study	Group or Individual Randomization?	Sample Sizes in Treatment (T) and Control (C) Groups	C.V. Profits	C.V. Revenues	Attendance Rate	Power to Detect Increase of:			
						25% in Profits	50% in Profits	25% in Revenues	50% in Revenues
Berge et al. (2011)	Group	119 (T), 116 (C) groups (a)	0.80	1.47	83%	0.631-0.842	0.996-1.000	0.239-0.365	0.705-0.897
Bruhn and Zia (2012)	Individual	297 (T), 148 (C)	2.69	n.a.	39%	0.070	0.132	n.a.	n.a.
De Mel et al. (2012)	Individual	200 (T1), 200 (T2), 228 (C)	0.49	0.91	70%	0.990	1.000	0.632	0.994
Drexler et al. (2012)	Individual	402 (T1), 404 (T2), 387 (C)	n.a.	1.63	49%	n.a.	n.a.	0.231	0.686
Giné and Mansuri (2011)	Group	373 (T), 374 (C) groups	n.a.	n.a.	50%				
Karlan and Valdivia (2011)	Group	138 (T), 101 (C) groups	-24.96	2.30	80%	0.057 (b)	0.078 (b)	0.120-0.757	0.335-1.000
Klinger and SchündelIn (2011)	Individual RD	377 (T), 278 (C)	n.a.	2.51	n.a.	n.a.	n.a.	0.259 (d)	0.746 (d)
Mano et al. (2012)	Individual	47 (T), 66 (C) (b)	1.23	1.20	87%	0.188	0.571	0.195	0.592
Sonobe et al. (2011)									
Tanzania	Individual	53 (T), 59 (C)	1.99	1.61	92%	0.109	0.292	0.141	0.414
Ethiopia	Individual	56 (T), 47 (C)	1.94	2.49	75%	0.087	0.204	0.072	0.142
Vietnam - Steel	Individual	110 (T), 70 (C)	1.59	0.93	39%	0.075	0.153	0.124	0.353
Vietnam - Knitwear	Individual	91 (T), 70 (C)	-8.15	2.34	59%	0.052	0.058	0.074	0.150
Valdivia (2011)	Individual	709 (T1), 709 (T2), 565 (C)	n.a.	2.29	51%	n.a.	n.a.	0.207	0.626

Notes:

n.a. denotes not available, either because the study did not report this outcome, or because it didn't report the coefficient of variation (C.V.).

Personal correspondence with authors used to obtain C.V.s from studies which only report sample means and not standard deviations.

Where range is shown, first number is power if intra-cluster correlation is one, second is power if intra-cluster correlation is zero.

(a) Numbers in control and training only groups - the study also includes groups with grants.

(b) Power calculation assuming randomization was at the individual level. Actual power will be lower once group-level randomization is accounted for.

(d) Study does not examine revenue as an outcome, since some data is collected retrospectively.

Power calculations ignore survey attrition, which would further lower power. They also assume entire sample is an existing enterprise.

Attendance rate for Klinger and SchündelIn (2011) assumed to be 90% for purpose of power calculations.

Power calculations assume one baseline and one post-treatment survey, with an autocorrelation in the outcome variable of 0.5, and ANCOVA estimation.

**Table 6: Follow-up Survey Timing of Different Studies**

Study	Number of Follow-up Surveys	Months since Intervention	Attrition rate
Berge et al. (2011)	1 (b)	5 to 7	18
Bruhn and Zia (2012)	1	5 to 6	11
De Mel et al. (2012)	4	4, 8, 16, 25	6 to 8
Drexler et al. (2012)	1	12	13 to 46 (a)
Field et al. (2010)	1	4	5.3
Giné and Mansuri (2011)	1	19-22	16
Karlan and Valdivia (2011)	1	12 to 24	24
Klinger and Schündeln (2011)	1	12	28
Mano et al. (2012)	1	12	17
Premand et al. (2012)	1	9 to 12	7.2
Valdivia (2011)	1	7 to 10	18

Notes:

(a) Attrition rate is 46% for business outcomes like sales, 13% for business practices.

(b) A second follow-up survey was conducted 2 years after training, but results are not yet available from this.

**Table 7: Impacts of Business Training on Business Start-up and Survival**

Study	Gender	Impact on Survival		Impact on Start-up		
		Point estimate	95% CI	Point estimate	95% CI	
Bruhn and Zia (2012)	Mixed	0.013	(-0.09, +0.10)	0	n.r.	
	Female	-0.125	n.r., not sig.	0	n.r.	
	Male	0.072	(-0.07, 0.21)	0	n.r.	
De Mel et al. (2012)	Current Enterprises	Female	-0.026	(-0.102, +0.051)	n.r.	n.r.
	Potential Enterprises	Female	n.r.	n.r.	<b>+0.09 (4 months)</b>	<b>(0, 0.18)</b>
Giné and Mansuri (2011)	Mixed	0.034	(-0.021, 0.089)	-0.006	(-0.11, 0.07)	
	Male	<b>0.061</b>	<b>(-0.012, 0.133)</b>	-0.011	(-0.04, +0.01)	
	Female	0.001	n.r., not sig.	0.002	n.r., not sig.	
Karlan and Valdivia (2011)	Female	n.r.	n.r.	-0.019	(-0.05, +0.01)	
Klinger and Schündeln (2011)	selected in first phase	Mixed	n.r.	n.r.	0.044	(-0.12, 0.21)
	selected in first phase	Female	n.r.	n.r.	-0.019	(-0.31, +0.27)
	getting trained in second phase	Mixed	n.r.	n.r.	<b>0.465</b>	<b>(0.10, 0.82)</b>
		Female	n.r.	n.r.	<b>0.572</b>	<b>(0.04, 1.10)</b>
Mano et al. (2012)	Male	<b>0.095</b>	<b>(0.022, 0.167)</b>	n.r.	n.r.	
Premand et al. (2012)	Mixed	n.r.	n.r.	<b>0.04</b>	<b>(0.02, 0.06)</b>	
	Male	n.r.	n.r.	<b>0.06</b>	<b>(0.04, 0.08)</b>	
	Female	n.r.	n.r.	<b>0.03</b>	<b>(0.01, 0.05)</b>	
Valdivia (2011)	General training	Female	<b>-0.035</b>	<b>(-0.074, +0.004)</b>	0.017	(-0.02, +0.05)
	Training + technical assistance	Female	-0.009	(-0.034, +0.016)	-0.013	(-0.04, +0.03)

Notes:

95% CI denotes 95 percent confidence interval. Impacts significant at the 10% level or more reported in **bold**.

n.r. denotes not reported. Not sig. denotes point estimate is not significantly different from zero.

Berge et al. (2011) and Drexler et al. (2012) do not report impacts on either survivorship or start-up

**Table 8: Impact of Business Training on Business Practices**

Study	Gender	Units	Number		95% CI
			of Practices	Point estimate	
Berge et al. (2011)	Male	p.p.	n.r. (a)	0.13- <b>0.26</b>	n.a.
	Female	p.p.	n.r. (a)	0.08-0.17	n.a.
Bruhn and Zia (2012)	Mixed	s.d.	3	<b>0.272</b>	<b>(0.03, +0.51)</b>
	Male	s.d.	3	<b>0.290</b>	<b>(0.01, 0.57)</b>
	Female	s.d.	3	0.214	n.r.
De Mel et al. (2012)					
Current Enterprises	Female	num	29	<b>2.03</b>	<b>(1.27, 3.30)</b>
Potential Enterprises	Female	num	29	0.87	(-0.23, +1.97)
Drexler et al. (2012)					
"Rule-of-thumb"	Mostly Female	s.d.	12	<b>0.14</b>	<b>(0.06, 0.22)</b>
"Standard"	Mostly Female	s.d.	12	0.07	(-0.03, 0.17)
Giné and Mansuri (2011)	Mixed	s.d.	3	<b>0.131</b>	<b>(0.01, 0.25)</b>
	Male	s.d.	3	0.114	(-0.05, 0.28)
	Female	s.d.	3	0.140	n.r.
Karlan and Valdivia (2011)	Mostly Female	s.d.	14	<b>0.03</b>	<b>(0.00, 0.06)</b>
Mano et al. (2012)	Male	p.p.	3	<b>0.24-0.42</b>	n.a.
Valdivia (2011)					
General training	Female	s.d.	11	-0.018	(-0.04, +0.01)
Training + technical assistance	Female	s.d.	11	<b>0.049</b>	<b>(0.02, 0.08)</b>

*Notes:*

95% CI denotes 95 percent confidence interval. Impacts significant at the 10% level or more reported in **bold**.

Units for measuring practices are either standard deviations of a normalized aggregate (s.d.), percentage points (p.p.), or number of distinct practices improved (num).

Number of practices is the total number of practices measured.

When no aggregate measure is reported, the range of point estimates for individual practices is given.

n.r. denotes not reported. n.a. denotes not applicable since range of estimates given.

(a) three main practices reported, but one of them is a marketing index, and it is unclear how many underlying practices this index encompasses.

**Table 9: Impacts on Business Profits and Sales**

Study	Gender	% increase	95% CI	% increase	95% CI	
Berge et al. (2011)	Male	<b>23.5%</b>	<b>(-2%, +39%)</b>	<b>29.3%</b>	<b>(+1.6%, +64.6%)</b>	
	Female	-4.3%	n.r.	-0.5%	n.r.	
Bruhn and Zia (2012)	Mixed	-15%	(-62%, +32%)	n.r.	n.r.	
De Mel et al. (2012)	Current Enterprises	Female	-5.4%	(-44%, +33%)	-14.1%	(-68%, +40%)
	Potential Enterprises	Female	<b>43%</b>	<b>(+6%, +80%)</b>	<b>40.9%</b>	<b>(-6%, +87%)</b>
Drexler et al. (2012)	"Standard"	Mostly Female	n.r.	n.r.	-6.7%	(-24.5%, +11.2%)
	"Rule-of-thumb"	Mostly Female	n.r.	n.r.	6.5%	(-11.4%, +24.4%)
Giné and Mansuri (2011)	Mixed	-11.4%	(-33%, +17%)	-2.3%	(-15%, +13%)	
	Male	-4.3%	(-34%, +38%)	4.8%	(-14%, +27%)	
	Female	n.r. (a)	n.r.	n.r. (a)	n.r.	
Karlan and Valdivia (2011)	Mostly Female	17% (b)	(-25%, +59%)	1.9%	(-9.8%, +15.1%)	
Mano et al. (2012)	Male	54%	(-47%, +82%)	22.7%	(-31%, +76%)	
Valdivia (2011)	General training	Female	n.r.	n.r.	-9%	(-23%, +9%)
	Training + technical assistance	Female	n.r.	n.r.	<b>19.2%</b>	<b>(-1%, 44%)</b>

**Notes:**

95% CI denotes 95 percent confidence interval. Impacts significant at the 10% level or more reported in **bold**.

n.r. denotes not reported.

(a) They look at an aggregate sales and profitability measure and find no significant impact for either gender.

(b) Impact on profit from main product.

Profit increases are scaled as a percentage of the control group mean to enable comparability.