Managers Training

Authors: Matias Busso, Nicolás Irazoque, and Kyunglin Park

In this note, we describe the procedure used to search, filter and code information from studies that credibly measure the impact of managerial skills training on several specific outcomes (managerial skills, productivity, profits, firm survival).

SEARCH PROTOCOL

We started our initial search through Google Scholar and two key literature review papers (Hogendoorn, 2019; Mckenzie et al., 2014), which yielded a total of 7,012 papers. This group of studies contained "Management training," "Business training," or "Entrepreneurship training" in their titles or abstract. While we did not limit our search protocol by regions or their level of development, we did limit our search to studies conducted after the year 2000 and to those published in English.

In order to include only studies conducted in the business management field, we clarified the subject of training in the next set of keywords; "Firms," "Business," "Workers," and "Employees." Based on the 363 studies that included the aforementioned subject keywords, we then dropped duplicates, along with studies that focused largely on undertaking a systematic review of training. We then filtered studies to keep only those that measured the impact of managerial skills training by using the keywords "Effects," "Impact," "Effectiveness," "Return," and "Evidence" as indicative of studies that report on the returns of such training. **As a result, the search protocol has yielded a total of 176 papers for further review based on our inclusion criteria.**

INCLUSION PROTOCOL

Once we gathered a pool of papers, we developed a set of inclusion criteria based on the content of the intervention and the methodology of estimating the effects. We focused on papers in which the main objective of the intervention was to train managers or entrepreneurs to improve certain business management skills and hence improve firm performance. Therefore, we did not consider microfinance or other types of financial aid interventions, including cases where role models were selected from peer participants to provide a good example. If the intervention had both a financial assistance component and a consulting component that could not be separated, it was included in the analysis. However, if studies conducted multiple treatments and the components could be separated, only the consulting services were included. We did not restrict the method of delivering the treatments, such as classroom training or on-site consultations.

We only included studies that had a control group from which researchers could credibly build a counterfactual in order to provide estimations of treatment effects of the intervention. We excluded papers reporting results using only propensity score matching methods or those using regressions without any underlying credible research design. We included studies that used one of the following: Randomized Control Trial (RCT), Regression Discontinuity Design (RDD), Difference-in-Difference or Instrumental Variables (IV). The specific requirements are detailed in Table 1.

∴ TABLE 1. Specific requirements about methods included

Method	Specific requirements
RCT	It must include a follow-up sample for the treatment group not targeted to compliers.
RDD	Study shows balance in covariates.
	Graph shows no jump of the density distribution for the forcing variable at the threshold.
	Includes convincing argument indicating that procedure to assign units is not prone to manipulation.
Difference-in-difference	Diff-in-diff specification or controlling for baseline outcome.
Instrumental variables	F-test for first stage above 10.
	Convincing argument of exogeneity of instrument.

CODING

Coding information reported from each study was based on two levels: evaluation (intervention) and effects. To code details on evaluation, we developed variables that describe the details of the intervention. They include general context summary of the training, how the treatment was designed, and the duration of the training took place, among other features. To code effects, or outcomes, we coded a set of variables that would provide specific parameters of each result. For instance, the variables included outcome name, the unit of measurement, how many months reporting the outcome was delayed, estimation coefficients, and number of observations.

After we have coded all outcomes by evaluation, we reviewed each effect and categorized them into four main groups and twelve sub-groups:

  i. **Profits:** profits, revenue, other

  ii. **Productivity:** productivity, size (measured by labor), size (measured in capital)

  iii. **Survival measures:** firm entry, firm survival, other

  iv. **Management skills:** operation practices, record-keeping, soft skills
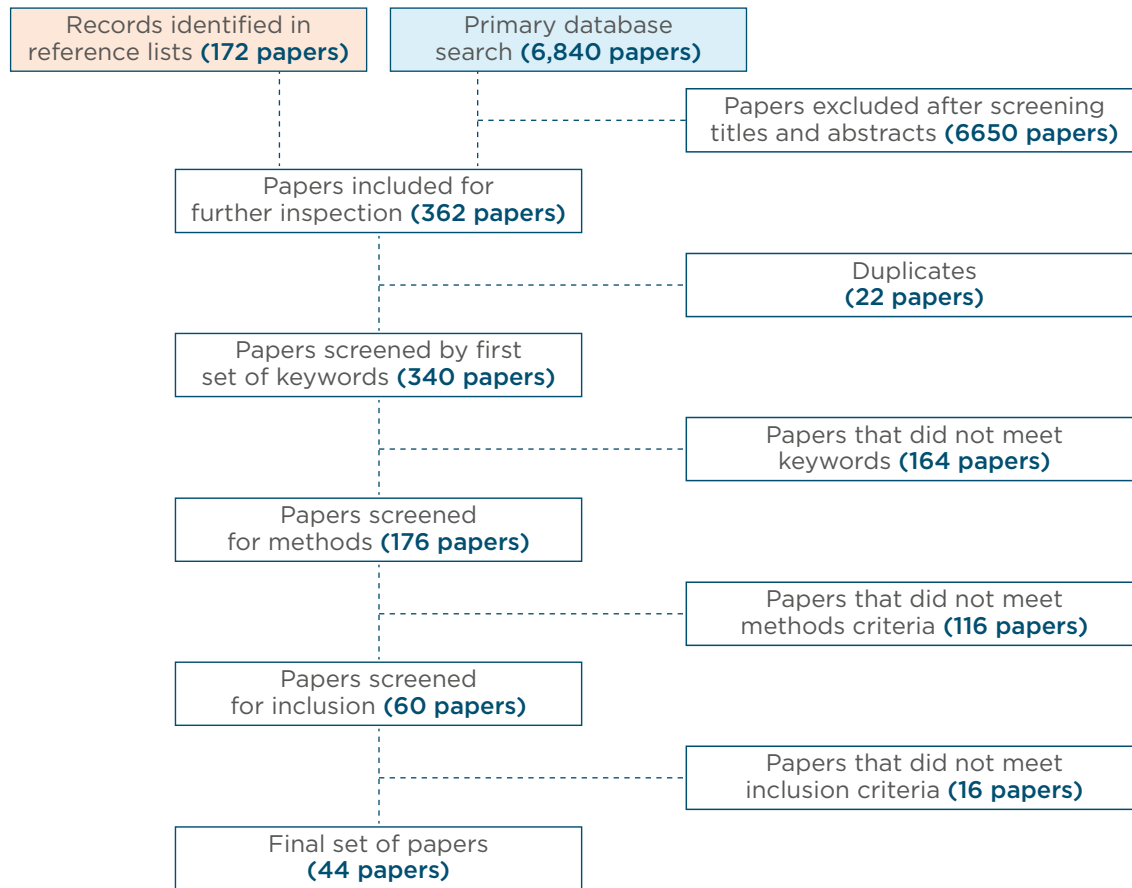
CALCULATING EFFECT SIZES

The majority of the studies included use experimental methods and a few quasi-experimental methods to estimate intention-to-treat effects or average treatment effects. The effect sizes were either computed by using $\hat{\beta}$, the parameter from a regression, or through the difference in means.

The unit of our analysis was the "evaluation." Due to the variation in the type of outcome each study report and its measurement unit, we needed a way to harmonize the effect sizes. In many cases studies also conducted multiple treatments, reporting several outcomes with multiple follow-up results. While we computed the "effect size" as $\hat{\beta}$ directly for estimating log change in profits and percentage change in firm survival, we used $\hat{\beta}/y_{c\,post}$ for cases where outcomes were reported in profit levels or percentage point changes. If the study did not report the post control group mean, post treatment, we alternatively used mean values from baseline.

For outcomes related to productivity levels and management skills, we use the Standardized Mean Difference (SMD): the effect size divided by its standard deviation. Following Kluve (2017), the SMD d can be calculated from the quotient of the regression coefficient $\hat{\beta}$ and the standard deviation of the effect $\hat{\sigma}$ (not the standard error)

$$d = \frac{\hat{\beta}}{\hat{\sigma}}$$

FIGURE 1. Stages and papers checked

Records identified in reference lists **(172 papers)**	Primary database search **(6,840 papers)**

Papers excluded after screening titles and abstracts **(6650 papers)**

Papers included for further inspection **(362 papers)**

Duplicates **(22 papers)**

Papers screened by first set of keywords **(340 papers)**

Papers that did not meet keywords **(164 papers)**

Papers screened for methods **(176 papers)**

Papers that did not meet methods criteria **(116 papers)**

Papers screened for inclusion **(60 papers)**

Papers that did not meet inclusion criteria **(16 papers)**

Final set of papers **(44 papers)**

We approximated $\hat{\sigma}$ as S_p from the standard error of the coefficient $\hat{\beta}$.

$$S_p = SE \times \sqrt{\frac{n_c \times n_t}{n_c + n_t}}$$

We use the estimated standard error from the regressions as SE. The variance of $d(V_d)$ depends on the SMD (d) itself, the t-statistic t and the degrees of freedom of the regression coefficient, which is in general $v=n-k$. In this case we stated $k=2$ because it is the minimum number of coefficients estimated. d and Vd are not sensitive to the value of k in our sensitivity analysis.

$$V_d = \frac{d^2}{v-2}\left(\frac{v}{t^2} + v\,c^2\,(v) - v + 2\right)$$

where $\quad \frac{1}{c(v)} = \sqrt{\frac{v}{2}}\frac{\Gamma\left(\frac{v}{2}-0.5\right)}{\Gamma\left(\frac{v}{2}\right)}\quad$ and $\Gamma()$ is the gamma function. When the study has a large n, $c(v)=1$.

Some papers report the confidence interval, in which case we used the formula $IC=\hat{\beta}\pm1.96\times SE$ to recover SE. When the paper reported the t-stat we used the formula

$$\frac{\hat{\beta}}{SE} = t$$

In order to summarize different outcomes, during the coding phase we selected the most aggregated effect. For example, if there are effects for male, female, and pooled sample we only use the pooled sample. Given that business performance may be difficult to measure, we found a wide range of different outcomes measuring entrepreneurial quality. Therefore, after we listed the outcomes related to management quality, we reviewed case by case to identify commonly reported outcomes and to group them into similar categories.

Additionally, we combined effects across time if the study conducted more than one follow-up survey. Initially, if there were more than two time periods, we used the earliest and latest time period. Then, we followed Borenstein et al. (2019) to combine effects and their standard errors assuming a correlation of 0.5 for standard error computation.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

September 2022