

Review of “An Impact Evaluation of the MIDA FBO Training”

Alan de Brauw, Markets, Trade, and Institutions Division, International Food Policy Research Institute

Impacts of the Project

According to the report, the MIDA/FBO agricultural project was “designed to enhance the profitability of staple food and horticultural crops and to improve delivery of business and technical services to support the expansion of commercial agriculture among farmer-based organizations (FBOs). Funds from the MCC compact were used for farmer and enterprise training; irrigation development; land tenure facilitation; improvement of post-harvest crop handling and value chain services; improvement of credit services; and the rehabilitation of feeder roads.

To evaluate the MIDA/FBO training project, a randomized control trial design was used. The evaluation took place in two waves or batches, and each wave included about 600 FBOs. Within each FBO, 5 farmers were sampled, which would imply that the sample size should be about 3000 for each “batch” per round. This sample design was deemed adequate to identify changes in profits or crop productivity of around 10 percent. Because the batches could not be kept in the control group for a long period of time (as the project was being phased in), the decision was made to conduct the follow-up surveys one year after the baseline, so that the intervention would have just taken place.

Clearly, the project was quite large in scope, and taking a fresh look at the project, it is important to rethink the outcomes on which we might expect to see impacts. Specific farmers or even specific FBOs might be affected by part of the project, but not all of the project. For example, irrigation development and the rehabilitation of feeder roads almost certainly did not take place everywhere. Further, one would expect that the trainings might have had a heterogeneous effect on farmers for two reasons. First, it is clearly voluntary whether or not farmers actually attended the trainings. Since much of the intervention came in the form of trainings, it could be that specific farmers or groups of farmers simply did not attend many of the trainings, and as a result we would expect weak evidence of impacts. Second, it could be that farmers knew a great deal of what they were taught already; if so, we would not expect the trainings to have changed behavior.

The sample design may work against finding impacts from an intent to treat perspective, for the following reasons. Consider that an important parameter in whether or not the intervention is effective would be the participation rate in trainings. Within the FBO, the estimated participation rate is quite noisy: the point estimate is 0, 20, 40, 60, 80, or 100 percent (with zero attrition), and the standard error on that point estimate is quite high if it is not zero or 100 (and so it didn't vary). If participation in trainings was low; e.g. 60 percent, then there is a reasonable probability that the point estimate is actually zero in a large number of FBOs. Further, contamination of the control group can also become a larger concern, and that seems to be the main problem in this case. Note that the increase in farmers reporting that they received information from MiDA training services is actually larger in the “control” group than in the treatment group in both batches (Table 8-2 in the descriptive report). So a third source of difficulties finding impacts could be that the training leaked into the control group, or quite clearly began in the control group before the evaluation took place.

So it can safely be concluded that farmers in the control group had approximately similar access to trainings, at least by the round 2 surveys, to nullify any potentially estimated impacts of those trainings on agricultural outcomes. Without an impact of trainings, there are really only three or four sources of potential impacts remaining as part of the program: potential improvements in irrigation development; rehabilitation of feeder roads; any additional finance available for farming; and the starter pack. As previously mentioned, it is unlikely that the irrigation development or feeder roads would have had much of an impact (and furthermore, they were incomplete by the time the evaluation was completed). So then we are left with the starter pack as the main source of potential project impacts.

That said, there are some points that look reasonably good in the descriptive statistics; for example, in Table 4-11 there are potentially some interesting differences in man-hours spent. I would suggest adding standard errors (not standard deviations) clustered at the FBO level to this table (and all of the descriptive tables) so that a reader can at least eyeball whether or not the differences appear statistically significant.

Not surprisingly, the report finds some small impacts that are attributed to the starter pack and/or finance. Specifically, it finds that farmers appear to take out additional loans from MiDA as a result of the program; it also finds increases in the use of chemicals and the value of seeds used, that are not quite as large as the value of the starter pack.

The question remains: were there impacts of the program that have not been measured? There are several aspects of the impact estimates that are somewhat curious and might deserve a second look before firm conclusions are made about this report. Before I get to statistical reasoning, it is probably not correct to call the impacts “the impact of training,” given that Table 8-2 in the descriptive report suggests that the control group received as much training from MiDA by the round 2 surveys, if not more, than the treatment group. This table is simplistic—it does not measure the level of exposure to training—but it

From a statistical perspective, one major concern I have with the bulk of the impact estimates is that there are clearly conditions placed on a number of the samples, which muddles the interpretation of some of the statistics. We would like to be able to interpret the “treattime” variable as the average treatment effect on those who were intended to be treated (the intent to treat effect). But it seems like the measure is often conditional on something, which affects the interpretation of the estimates. Take Table 6-2 for example, which is examining “income” of farm households (is this crop income?). There are only 1756 total observations in these regressions, whereas there were 10,692 in the previous table (which sounds like the total sample size, particularly given the attrition rate). Other tables are similar; why are there only 1659 observations in Table 6-5? Or what are these estimates conditional on? Literally none of the rest of the estimates use a sample of 10,692, with the exception of Table 6-12, which uses 26,606 observations (and must allow for different crop types, perhaps, within households)?

In summary, the impact results are totally inconclusive. They do appear to find an impact of the starter pack, but potentially the contamination of the control group affected the results here, as the control group reports receiving information on farming from the MiDA. Second, the apparent conditions placed on many of the estimates muddle their interpretation. Third, it is worth again noting that the impacts were measured after only a one year period. If this

evaluation could have taken place over a longer period of time, it is possible that impacts could have been larger.

Objectives and Discussion

The objective of this project was to increase farmer incomes so that there was a measurable rate of return on MCC investments. Clearly, since impact estimates are inconclusive at best we cannot conclude that the project was successful in obtaining that objective, at least in the immediate term. It might be worthwhile learning now about how agricultural incomes and opportunities have changed over time, though with the understanding that those changes cannot necessarily be attributed to the program. Clearly, an important part of the benefits of a training program could be increased outcomes for a long period of time, which helps the benefits outweigh the costs. With only one year of measured impacts (which are largely zero), it is quite difficult to understand what the impacts of this project really were. The true impacts of agricultural projects likely take time to grow; note that the feeder road interventions were not yet complete when the evaluation took place.

I have several recommendations about the design and implementation of the evaluation that might be useful for future MCC/MCA decision making. Most of my recommendations deal with the evaluation component, but some deal with the implementation and design of projects as well. First, it seems to me that more thought related to intermediate outcomes in the project could have informed a richer impact evaluation. If the evaluation surveys had incorporated carefully designed knowledge tests or survey questions both before and after the trainings, we might have learned about what knowledge from the trainings were picked up by farmers, and which were not. The questionnaire might have also asked directly about participation in trainings; but even better would be for the project to include simple monitoring forms so that the evaluator can know who received training and who did not. Obviously, such knowledge abstracts from the rationale for randomization and a clean treatment, but it is somewhat obvious to me that this program was quite heterogeneous, and learning more about correlations between trainings attended and outcomes might have shed light on more of the potential impacts (or efficacy) of this particular project. I should note that such data need not come from the survey questionnaire; it could also potentially come from monitoring data collected during the course of the intervention.

Second, there is a clear risk in planning evaluations over very short time horizons, such as one year. In this case, it might have made more sense to use larger samples of farmers in FBOs that entered quite early and quite late, so that two year impacts could have been estimated. Obviously, it is difficult to plan in advance for such time horizons, but in this case it seems it could have helped the evaluation; we have no idea what the two year impacts are.

Third, it is worthwhile thinking through the “theory of change” for specific variables that are often measured in agricultural evaluations. What has to happen, in terms of producer behavior, for trainings to increase agricultural profits, for example? Profits and yields might respond quite differently to different components of the treatment. Although I am sure that this thinking took place in advance of this project and evaluation, it is somewhat lost in the report. As a result, the report feels a bit more like a fishing expedition (fishing for statistically significant impacts) than it might if we considered more carefully the reasons we might expect some variables to have changed.

Finally, from a program perspective, it is also worth thinking through the theory of change, and what has to happen from a behavioral perspective for projected rates of return to materialize. Training programs have a mixed record in the literature, in both agricultural and health/nutrition contexts, when impact evaluations are rigorous. It seems to me like incentives might not have been aligned properly for farmers to actively participate in this project, though I have no evidence to back up that statement. It would have been useful to know if specific parts of their knowledge base changed—then we would know if the reason impacts were not measurably different from zero in most cases arose from the lack of a change in knowledge or a problem translating knowledge into behavior. This type of knowledge is particularly interesting, as then one could conclude something about the types of programs that are necessary in the future—either programs that better transmit knowledge, or programs that help farmers translate knowledge into actions.