



## Impact Evaluation Research Design Recommendations

Final Report

12/11/2015

Mine Ekinci | Florencia Gay

EC.788 MIT D-Lab: Field Research

Instructor: Elizabeth Hoffecker-Moreno

## CONTENTS

1. Preface
2. Introduction
  - 2.1. Mission
  - 2.2. Core services/products
  - 2.3. Business model
  - 2.4. Organizational structure
  - 2.5. Theory of change
3. Impact measurement challenge
  - 3.1. Purpose of impact research
  - 3.2. Research questions
  - 3.3. Resources and constraints
4. Recommendations
  - 4.1. Income, production and behavior
    - i.* Research design
    - ii.* Data collection methods
    - iii.* Sampling and comparison groups
    - iv.* Timing
    - v.* Key findings from the evidence review
  - 4.2. Soil health
    - i.* Research design
    - ii.* Data collection methods
    - iii.* Sampling and comparison groups
    - iv.* Timing
    - v.* Key findings from the evidence review
  - 4.3. Satisfaction
    - i.* Research design
    - ii.* Data collection methods
    - iii.* Sampling and comparison groups
    - iv.* Timing
  - 4.4. Trainings
    - i.* Research design
    - ii.* Data collection methods
    - iii.* Sampling and comparison groups
    - iv.* Timing
    - v.* Key findings from the evidence review
5. References

## **1. PREFACE**

This report is prepared for Zasaka to provide them with research design recommendations for their impact evaluation research. The report consists of three main parts: Introduction, Impact Measurement Challenge and Recommendations.

In the Introduction part, there is a short summary of all the information that we collected about Zasaka from the interviews we made with the co-founders Carl Jensen and Sunday Silungwe and the documents we reviewed. At the end of the Introduction, we also included a revised version of Zasaka's Theory of Change that guided the rest of our work.

In the Impact Measurement Challenge part, we first explained why Zasaka needs an impact evaluation research and how and by whom the research results are likely to be used. Later, we listed all the research questions that came up during our interviews by priority; we also categorized them into four different groups by subject. Lastly, we briefly explained which resources are currently available for a potential impact evaluation research study, and related to that, what kind of constraints need to be kept in mind while designing the study.

The Recommendations part is also divided into four sections based on the four categories we divided the research questions into. For each category, we recommend different research designs. In each section, we made specific recommendations about what type of research design can be followed, which indicators can be measured, which methods can be used for data collection, how samples can be chosen (for the treatment and comparison groups) and when the data should be collected. Furthermore, again in each section, we also summarized some of the key findings we learned from our evidence reviews that influenced our recommendations.

## 2. INTRODUCTION

Zasaka is a for-profit social enterprise located in Chipata, Zambia. Was founded in June 2014 by Carl Jensen and Sunday Silungwe.

### 2.1. Mission

Generate lasting income with small-scale farmers.

### 2.2. Core services/products

Services: Loan for seeds, market contracts, trainings and support by Private Extension Agents<sup>1</sup> (bi-weekly trainings, field visits, 1:1 advising, support for administrative work)

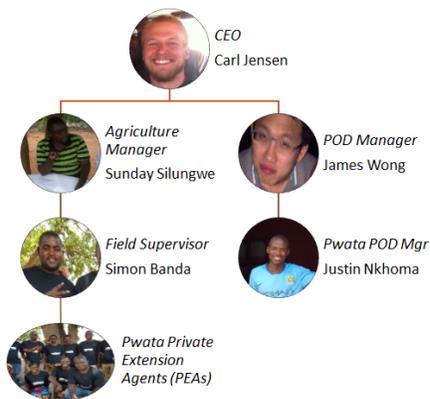
Products: Seed (maize, vegetables, and cowpea), manure, lime, rippers, hoes, chains for ploughing, herbicides, pesticides, fungicide, backpack sprayer, PICS bags

### 2.3. Business model

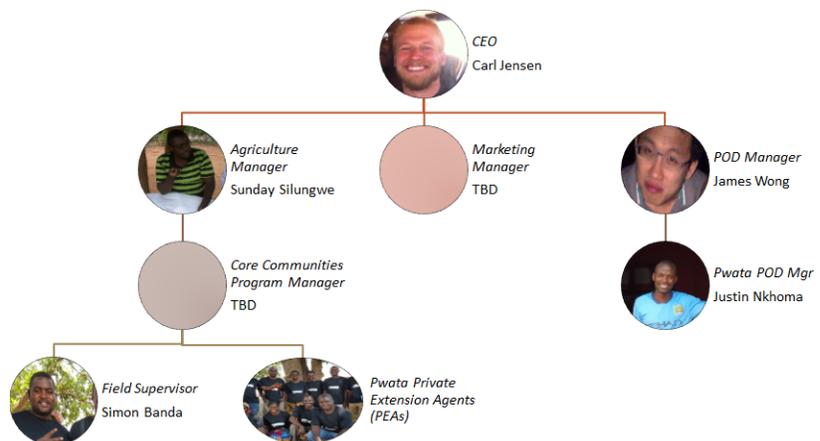
To date Zasaka has been chiefly financed by **grants**. However, starting from this year, they will shift their business model from one, which is grant-driven to one, which relies mainly on **impact investment**. Sales from the crops are another source of income.

Main lines of expenses are: 1- Inputs (e.g. seeds, fertilizers, pesticides), 2- Buy-back of crops from farmers. 3- Salaries.

### 2.4. Organizational structure



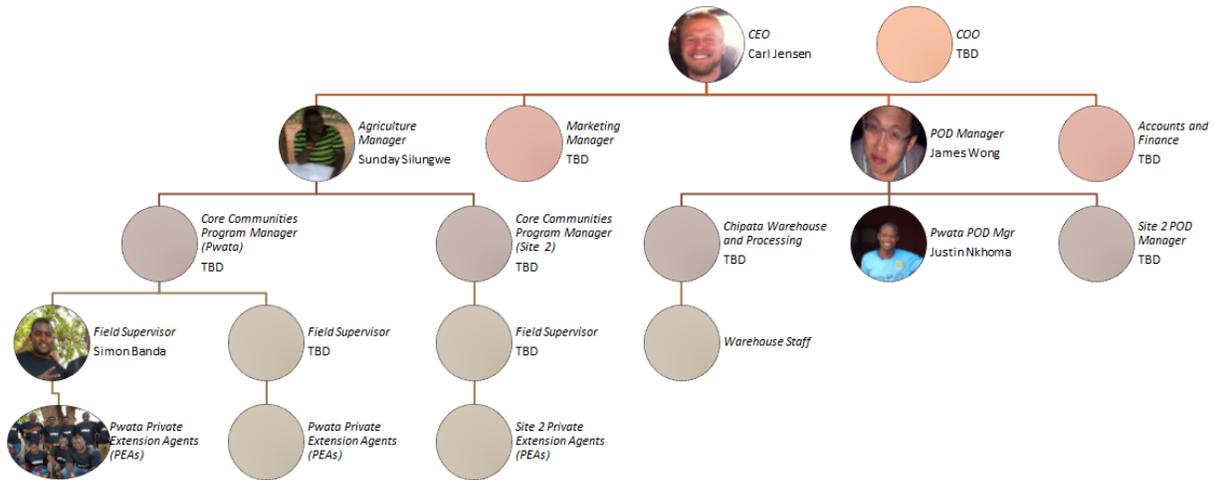
*Organizational structure.  
December 2015*



*Organizational structure.  
February 2016*

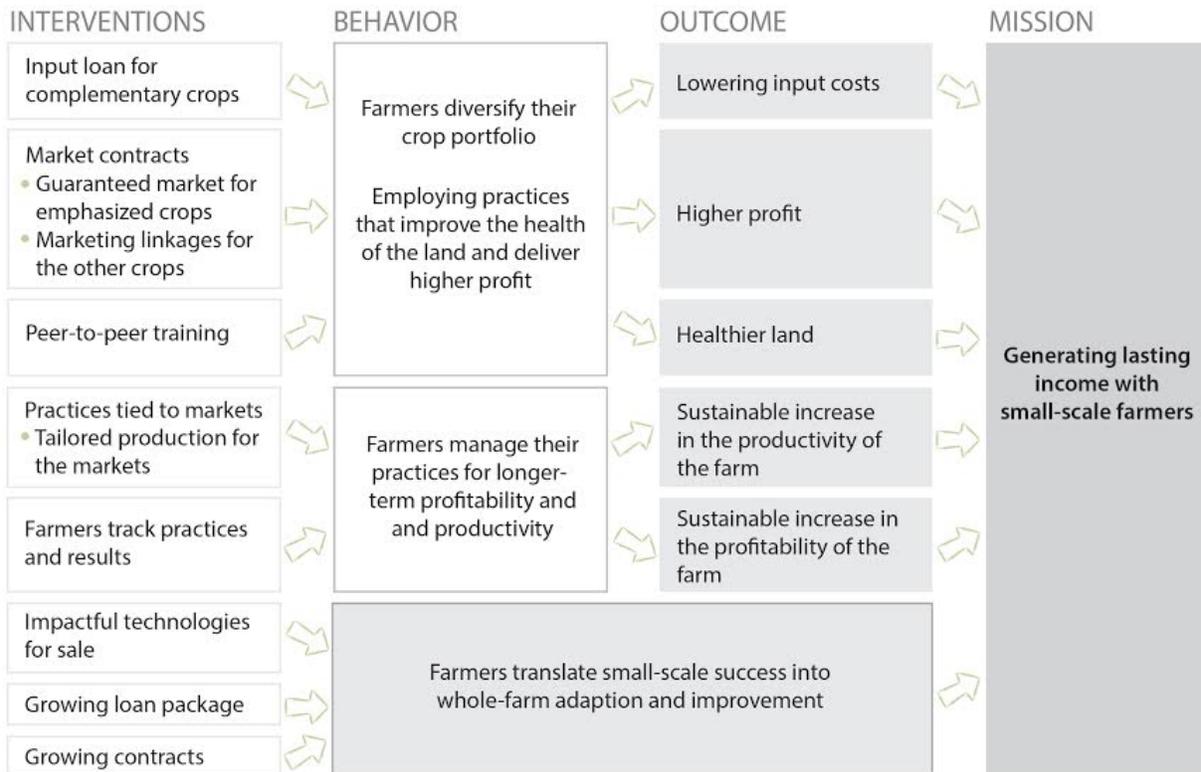
<sup>1</sup> Private Extension Agents (PEAs) are ZASAKA staff members working on the ground. PEAs are successful farmers that are chosen from ZASAKA farmers, they are all literate. PEAs act as the communicators between the organization of ZASAKA and the farmers. Each PEA is responsible of 40 farmers. Their primary responsibility is giving trainings and other types of support to other farmers to improve their practices.

<sup>2</sup> 310 is the sample size that will allow Zasaka to generalize the sample data to the population of Zasaka within 95%



Organizational structure. July 2016

## 2.5. Theory of change



### 3. IMPACT MEASUREMENT CHALLENGE

#### 3.1. Purpose of the impact evaluation research

Zasaka has three main purposes for conducting an impact evaluation research (ordered by descending priority):

1. To show ZASAKA's impact to existing and potential investors and donors,
2. To fix/redesign the parts of the program that are not working,
3. To show farmers the impact of ZASAKA, so:
  - a. They can improve their practices,
  - b. ZASAKA is perceived well by the farmers (important for the future of ZASAKA)

#### 3.2. Research questions

Below are the main research questions ordered by descending priority. We indicated how each research question relates to the theory of change that was delineated above. And we also grouped them under six main subjects: Income, Production, Behavior, Soil health, Satisfaction and Trainings.

- **Impact questions**

*Is the organization fulfilling its MISSION?*

Is the farmers' income rising? - **Income**

*Are the interventions leading to expected OUTCOMES?*

Are the farms that are engaged with Zasaka program more productive? – **Production**

Are the farms that are engaged with Zasaka program diversifying the crops they are growing? – **Production**

Is the health of lands that are used by Zasaka farms getting better/ worse or stay the same? – **Soil Health**

*Are the interventions leading to expected BEHAVIORS?*

Is there a change in the farmers' behaviors (loans, contracts, use of technology, various agricultural practices)? – **Behavior**

- **Key monitoring and evaluation questions that can be included in the research**

*Are the INTERVENTIONS implemented effectively?*

*How do the customers perceive the INTERVENTIONS?*

Are farmers satisfied with ZASAKA's activities? – **Satisfaction**

Are the trainings effective? – **Trainings**

- **Some further questions that can be integrated to the research**

*How do the customers perceive the INTERVENTIONS?*

Which activities and inputs (ground sheller, pesticides etc.) do the farmers perceived as useful? – **Satisfaction**

Are the farmers aware of all the options ZASAKA is offering them? – **Satisfaction**

How much do farmers appreciate ZASAKA's transparency? – **Satisfaction**

Are the farmers happier? – **Satisfaction**

### **3.3. Resources and constraints**

- **Staff and technical resources**

Mine (for two weeks) and Paula (for three weeks) are going to be in Chipata in January 2016 for collecting data. It may also be possible to get help from another MIT group who will be there in January. One of the co-founders, Sunday, is also likely to be available during this period to support the research team (Sunday has an international development research background, he has experience with data collection and analysis.)

After January 2016, the POD manager may be able to continue with data collection part-time.

Soil health measures are already being prepared and first measurements will be done in December 2015. Second round of measurements will be done in January 2016.

- **Financial resources**

Zasaka may be able to spend up to \$5,000 for the impact evaluation research.

- **Climate**

Tropical climate. January will be the rainy season, however it is still possible to reach the villages with no major problem.

- **Transportation**

2 cars and one motorbike are available for transportation.

- **Communication** (Telephone/Internet)

Wi-Fi connection is available in the office; data sticks can be used in the locations where there is telephone coverage (it is not constant in some places).

By December 2016, each PEA will be provided with Android phones that can be used for communication as well as collecting data.

- **Culture**

Separating women and men for the focus groups is advised.

People are approachable and open to talk in general.

- **Political situation**

Stable. Not relevant to the research.

#### 4. RECOMMENDATIONS

As explained above, we found that Zasaka's research questions are about different subjects and while some of them are strictly about impact, the others are more about monitoring and evaluation. Based on these categorizations, we think it is better for Zasaka to conduct multiple research studies rather than designing one single study to answer all questions. What we suggest is to conduct four different studies:

- 1- for measuring Zasaka's impact on **Income, Productivity and Behaviors**
- 2- for measuring Zasaka's impact on **Soil Health**
- 3- for measuring the **Satisfaction** of the farmers with Zasaka
- 4- for measuring the effectiveness of the **Trainings**

In this report, we give different recommendations for each of these areas. However, considering limited resources, it may also be a good idea for Zasaka to prioritize one or a few of these research studies over the others and conduct only these and leave the others to be implemented later. Nevertheless, in the long term, we think embedding monitoring and evaluation within the ordinary tasks of the organization may tremendously benefit Zasaka to prove their impact to the new donors and investors, to be able to have a better insight of the processes that leads to this impact, to be able to diagnose the problems and to fix them while they are scaling up. Furthermore, a good monitoring mechanism where the data is collected and analyzed with the farmers (this applies especially to the Income, Productivity and Behaviors part) can also work as an intervention in itself by enabling farmers to track their own practices and results in a systematic way.

##### 4.1. Income, production and behavior

RESEARCH QUESTIONS	INDICATORS
<b>INCOME:</b> Are the farmers' income rising?	Amount of crops sold Amount of crops consumed Price per kilo (calculated separately for each crop and may differ in different times of the year)
<b>PRODUCTION:</b> Are the farms that are engaged with Zasaka program more productive?  <b>PRODUCTION:</b> Are the farms that are engaged with Zasaka program diversifying the crops they are growing?	Number of seeds that are planted Types of seeds that are planted Area planted Pre-harvest yield in kilos (for each crop) Post-harvest yield (for each crop)
<b>BEHAVIOR:</b> Is there a change in the farmers'	Amount of lime that is used

behaviors (loans, contracts, use of technology, various agricultural practices)?	Is the liming done correctly? Amount of fertilizers that are used What types of fertilizers are used? Are the fertilizers used correctly? Amount of pesticides, fungicides, herbicides that are used? What types of pesticides, fungicides, herbicides that are used? Are the pesticides, fungicides, herbicides used correctly? Is the weed control being done correctly? Number of bags that were purchased for storing the crops
--	---

*i. Research design:* Longitudinal quasi-experimental research design with control groups.

We recommend a longitudinal quasi-experimental research design with comparison groups, first of all because such a design would provide Zasaka with a quite reliable proof of their impact. Second, even though longitudinal design is normally thought to be very costly, we think that Zasaka may be able to collect all this data by integrating data collection to the routinely tasks of the PEAs. This way, as PEAs are already working closely with the farmers on the ground, they are all literate and they will be provided with Android phones by December 2015, Zasaka may be able to collect rich data without necessarily hiring an extra staff.

Third, embedding a systematic monitoring process into PEAs and farmers' routinely work will allow PEAs and farmers to track together changes in farmers' incomes, productivity and agriculture-related practices. Since farmers' tracking of their practices and results is actually one of the strategies that Zasaka follows to fulfill its mission, we think that this monitoring component will enhance effectiveness of Zasaka's other activities and raise its eventual impact. Complementing the third component, fourth, is also known that in fact Zasaka is already working on creating a Farmers' Budget Workbook to track the production of the farmers. If such a Workbook can be extended to record some other indicators that are suggested above, it will be a great survey-like tool to collect data.

*ii. Data collection method:* Questionnaires (can also take the form of additional parts in the Farmers' Budget Workbook), observations and document review.

Questionnaires will be the main method for data collection; however, we think it is a good idea to triangulate the self-reported answers of the farmers with observations by PEAs (especially for measuring behavioral change) and document review (e.g. when available reviewing receipts of the crops sold together with the farmers before recording how much crops were sold).

*iii. Sampling and control groups*

Two options for choosing the samples for the treatment groups:

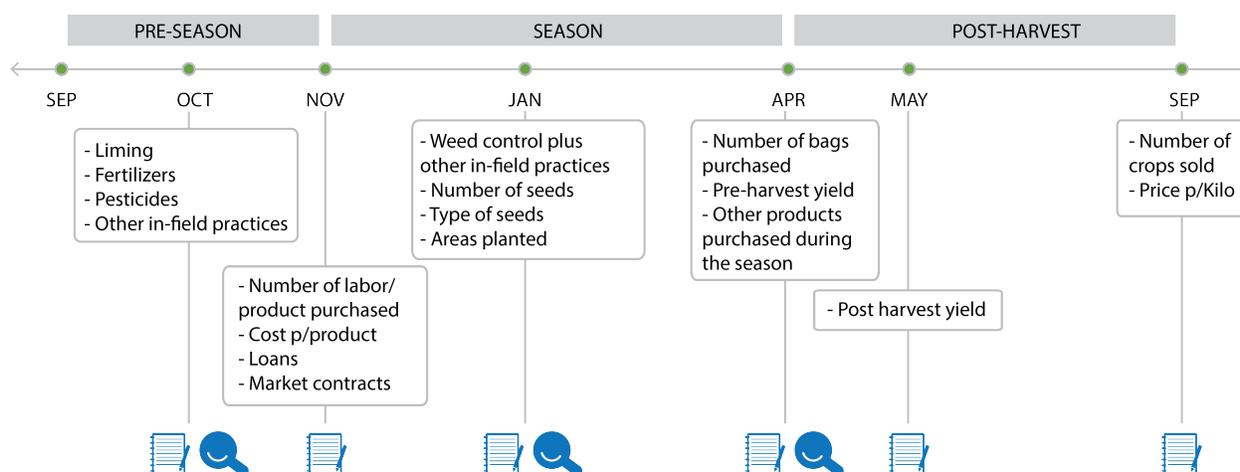
1- For the treatment group, collecting data from all Zasaka farmers, but later analyzing the monitoring data only from 310<sup>2</sup> randomly chosen the farmers.

2- For the treatment group collecting data only from 310 farmers and analyzing this data.

The control group will consist of randomly chosen farmers who will be Zasaka farmers the following year (*pipeline sampling*). However, it is also possible to create some additional comparison groups, if Zasaka wants to attribute the changes in income, production and behaviors to its specific activities. For example, if Zasaka is interested in knowing how much effect trainings have on the income of farmers, they could create an additional comparison group consisting of farmers who only received trainings from Zasaka but haven't used any of their products or other services. Comparing these comparison groups with the original control group would allow Zasaka to track the impact of their specific activities.

#### iv. Timing

Pre-test data collection in January 2016 for the to-be-Zasaka farmers in 2016-2017 season; data will continue to be collected every year from the same sample of farmers for at least three more years. In the mean time, each year the data will also be collected from to-be-Zasaka farmers (obviously, this group will be changing each year) since they will be the control group for the study. As measurement of all the indicators is time-sensitive, we recommend the data to be collected not once in a year but six times in a year in chunks. This is a sample timeline we recommend for collecting data to be repeated each year:



#### v. Key findings from the evidence review

##### **Another perspective to measuring income: Can consumption be a better indicator?**

In this report, we recommend measuring income three times a year and we measure it by multiplying the amount of crops sold and consumed with the average price that they were being/could be sold in the market. However, it needs to be recognized that this method only captures the agricultural income of the households. We think for the purposes of this research, this measurement would be enough to show Zasaka's impact on the farmers' incomes; however, if overall financial well-being of the farmers want

<sup>2</sup> 310 is the sample size that will allow Zasaka to generalize the sample data to the population of Zasaka within 95% confidence interval and with 5% margin of error.

to be captured, measuring the consumption of the farmers would be a better indicator:

- World Bank (2000) argues that “Consumption is conventionally viewed as the preferred welfare indicator, for practical reasons of reliability and because consumption is thought to better capture long-run welfare levels than current income.”
- Both income and consumption indicators are self reported through surveys and might contain errors. However, measurement errors in consumption are less pronounced for those with few resources than is measurement error in income (Meyer, 2003).
- It is common to distinguish four main components in the measurement of income: (i) wage income from labor services; (ii) rental income from the supply of land, capital, or other assets; (iii) self-employment income; and, (iv) current transfers from government or non-government agencies, or other households.
- Most surveys collect data on four main classes of consumption: (i) food items; (ii) non-food items; (iii) consumer durables; and, (iv) housing.
- Regarding to timing, a balance must be struck between the desire for minimizing recall and reporting error (which argues for a short reference period) and the desire to capture permanent rather than transitory living standards (which argues for a long reference period). For farm households, whose income is seasonal, a year is the only sensible reference period (Fields, 1993).

#### **An exemplary study for measuring productivity**

A research study that measured the rates of return to fertilizers in Kenya (Duflo, 2008) might be relevant to Zasaka’s research design. It involved a series of six field trials over three years designed to ascertain the profitability of fertilizers on farms and was implemented with the NGO International Child Support (ICS).

- Farmers were selected from lists of parents at local schools, and the ICS paid for fertilizer and hybrid seeds, delivered materials, helped these farmers apply fertilizer and seeds, and assisted them with the harvest.
- On each farm, an ICS field officer measured 3 adjacent plots. The first two would receive different treatments and the third one was a comparison plot kept directly next to treatment plots, which were farmed using traditional methods. The type of seed and amount of fertilizer applied to each plant was varied by plot, but farmers were instructed to tend all plots exactly the same. The differences between the treatments applied in the first two plots were decided based on common practices and recommendations of the Kenyan Ministry of Agriculture.
- By having interviews with the farmers and field observation during the seasons, ICS controlled that the treatment was being done properly.
- Timing: 6 growing seasons.

#### **Important points to consider while assessing behavioral change**

- Since behavioral change is a complex process and often it occurs through time; it is advised to use a mixture of qualitative and quantitative data. Especially, there may be some cases that no

quantitative proof can be found about the behavioral change, but with qualitative methods it can be understood that actually the intervention is working but there are some other factors that hinders the behavioral change or some more time is needed for the behavioral change to be reflected on the quantitative data (Sweedey, 2009).

- It is important to question if there are some environmental constraints that cause the intervention not to translate into behavioral changes (Raab, & Pasqual, 2011). In such cases, it may be interesting also to look at the behavioral intentions and try to specify the environmental constraints.

### Choosing control groups

While control groups are being chosen it is important to choose them from a location relatively distant to ensure there was no knowledge transfer from the trained farmers to the untrained (Tesfaye & Karpipai, 2010).

## 4.2. Soil health

We know that Zasaka already works on this part of the impact evaluation research, but the team will still make some recommendations based on the literature we reviewed that may be helpful for Zasaka to improve their existing research plan.

RESEARCH QUESTIONS	INDICATORS
Is the health of lands that are used by Zasaka farms getting better/ worse or stay the same?	Aggregate stability Available water capacity Bulk density Earthworms Infiltration Particulate organic matter Potentially nitrogen Reactive carbon Slaking Soil crusts Soil electrical conductivity Soil enzymes Soil nitrate Soil pH Soil respiration Soil structure and macropores Total organic carbon

*i. Research design:* Pre- and post- research design with a comparison group.

*ii. Data collection method:* Soil health tests

A standards rubric should be prepared beforehand for a meaningful interpretation of the test results.

### *iii. Sampling and control groups*

The fields in the comparison group should have similar characteristics to the treatment group at the pretest phase mainly in terms of their composition and humidity level (USDA, n.d.).

Samples should be taken from “representative areas” and “problem areas” in the field and they should be differently labeled. For each area, a few different samples should be taken. Later, when comparisons are made same categories of samples should be compared (e.g an inter-row representative sample should be compared with another inter-row representative sample).

### *vi. Timing*

Annual. After harvest or end of the growing season are the best times to make the measurements. To see the trend it is important to repeat the measurements between 3-5 years.

### *vii. Key findings from the evidence review*

#### **US Natural Resources Conservation Service and soil health tests**

US Natural Resources Conservation Service is the primary resource we used for designing the soil health component of our program.

Here is some more information we gathered on measuring soil health from USNRCS resources:

- It is suggested that all the tests should be applied instead of choosing a few due to the holistic nature of soil health (USDA, 2008).
- Measurements should be taken periodically over time to monitor changes and trends; annual sampling is particularly suggested. Timing of sampling matters; the best time to make measurements is considered to be when the climate is stable and “there have been no recent disturbances such as after harvest or the end of the growing season” (USDA, n.d.)
- Soil properties also vary across the field even though it is the same type of soil, different operations are also effective on the variability:

General field characteristics to consider are: row versus inter-row areas, differences in soil type, differences in management, wheel versus non-wheel tracked areas, differences in crop growth, salt affected versus non-salt affected areas, eroded versus non-eroded areas, differences in slope, and wet versus non-wet areas (drainage). (USDA, n.d.)

- It is important to select sample sites which are representative of the field as well as those that are representative of the trouble spots.

The number of samples or measurements to take will depend on the variability of the field. It is recommended that a minimum of three samples or measurements be collected on any one soil type and management combination. In general, the greater the variability of the field, the greater the number of measurements are needed to get a representative value at the field scale. When measuring EC, pH, and soil nitrates at the field scale, eight or nine sample cores from across the field could be bulked and mixed,

and two subsamples from the mixed cores could be analyzed. (USDA, n.d.)

### 4.3. Satisfaction

RESEARCH QUESTIONS	INDICATORS
Are farmers satisfied with ZASAKA’s activities?	Net Promoter Score: Promoters & Detractors
Which activities and inputs such as ground sheller, and pesticides, do the farmers perceive as useful?	How likely is it you would recommend Zasaka to a friend?
Are the farmers aware of all the options ZASAKA is offering them?	Why?
How much do farmers appreciate ZASAKA’s transparency?	
Are the farmers happier?	

*i. Research design:* Post-test analysis of the treatment group (no baseline data or control group)

*ii. Data collection methods:* Questionnaires and semi-structured interviews

We recommend using the net promoter system to identify promoters and detractors as a way to measure satisfaction of the farmers by tracking their loyalty. The first step is to conduct surveys with a close-ended question and an open-ended question: How likely is it you would recommend Zasaka to a friend? Why? The goal is to identify promoters and detractors, and their reasons to be excited about participating in Zasaka and what they believe cause them to feel that way.

Once identified the promoters and detractors, we recommend the following step of the Net Promoter System, which implies to contact detractors and conduct semi-structured interviews and make follow-up questions to understand the root cause and determine the real problem that needs attention. Net Promoter System companies often find it useful to think in terms of the “Five Whys” Total Quality Management approach, repeating “why” for each answer until they arrive at the root cause, the point at which an internal change would prevent the situation from happening again.

This step should be followed by what the system calls “Closing the loop” which means to find the root causes of an individual customer’s problem, and, whenever possible, “fixing” the situation for that customer.

*iii. Sampling and control groups*

We recommend that all Zasaka’s farmers should take the interviews, and only the ones identified as detractors should participate of the semi-structured interviews. It might be interesting to apply this system to measure satisfaction of Zasaka’s employees, including PEAs.

*iv. Timing:* To be done annually between post-harvest and land preparation periods.

*v. Key findings from the evidence review*

## The Net Promoter model

The Net Promoter model is an approach in which you use your Net Promoter score to drive improvements. The Net Promoter Score is calculated using the answer to a single question, using a 0-10 scale: How likely is it that you would recommend [brand] to a friend or colleague? This is called the Net Promoter Score question or the recommend question.

Respondents are grouped as follows:

**Promoters** (score 9-10) are loyal enthusiasts who will keep buying and refer others, fueling growth

**Passives** (score 7-8) are satisfied but unenthusiastic customers who are vulnerable to competitive offerings.

**Detractors** (score 0-6) are unhappy customers who can damage your brand and impede growth through negative word-of-mouth.

Subtracting the percentage of Detractors from the percentage of Promoters yields the Net Promoter Score, which can range from a low of -100 (if every customer is a Detractor) to a high of 100 (if every customer is a Promoter).

The Net Promoter System requires that the organization and its senior management commit to:

1. **Sorting customers** (and employees) into promoters, passives and detractors
2. Creating short-cycle **closed-loop** feedback, learning, recovery and action processes
3. Making it a **top priority to earn the enthusiastic loyalty of customers** (and employees)

The system itself comprises eight essential processes:

### 1. Reliable metric

### 2. Loyalty economics

### 3. Root cause

### 4. Closed loop

One of the most important elements of a meaningful Net Promoter system, closing the loop means immediately sharing customer feedback with employees, and responding directly to individual customers to take action on their concerns (and delights).

### 5. Learning

Short-cycle closed-loop feedback gives employees throughout the organization the opportunity to understand the impact of their actions on customers, which can enable faster learning. But learning requires more than feedback. It also demands deliberate processes to help employees experiment and get the sort of coaching they need.

### 6. Action

Beyond just collecting feedback, calculating scores and learning, Net Promoter companies actively seek ways to change day-to-day behavior throughout the organization. They have a bias toward action.

### 7. Robust operational infrastructure

### 8. Leadership and communication

Net Promoter companies require leaders who instill values in the organization and reinforce them through their actions, decisions and words. They demonstrate their own commitment to earning customer and employee loyalty, and demand it of others.

## Florence-Bain case

Florence, a not-for-profit social enterprise, launched a unique childcare model for working parents in

Tokyo in 2005. The service is non-facility based and involves dispatching experienced babysitters to look after children that fall ill. The model was popular, illustrated by 80 percent year-on-year growth since its launch. However, Florence was operating at a significant loss in its first two years and faced increasing financial difficulties due to highly unpredictable demand. Bain played a pivotal role in turning around the business, which broke even for the first time in 2009.

The Bain team developed a new pricing structure and demand forecast model. Bain also initiated quarterly surveys to track client and employee loyalty using the Net Promoter® score. The results: Florence turned profitable for the first time in 2009. That enabled Florence to expand services into Tokyo's suburbs and to open twelve new childcare centers. From 2011 to 2013, Bain periodically analyzed Florence Net Promoter® data to develop initiatives for improving customer and employee satisfaction (Bain, 2009).

#### 4.4. Trainings

RESEARCH QUESTIONS	INDICATORS
Are the trainings effective?	Farmers' level of satisfaction with the trainings Gain of knowledge Gain of skills Change in attitudes

*i. Research design:* Pre-, mid- and post- (with no comparison group)

*ii. Data collection methods:* Questionnaires (to measure satisfaction and change in attitudes) and tests to measure knowledge and skills.

Some sample questions for the pre-test questionnaires are: “What do you expect to learn from the trainings?” “Do you have any worries about the trainings? If yes, what worries you the most?” Some sample questions for the -mid and -post test are: “Which parts of the training you found the most beneficial?” “What can be improved about the trainings?”. We suggest that the questionnaires to be implemented by other PEAs (not by the ones who trained this particular set of trainees), while the other PEAs will know very well about the content and structure of the trainings, they can still be able to gather relatively more unbiased answers from the trainees.

For the tests, it is important to have a list of learning goals beforehand; to be clear which knowledge, attitudes and skills want to be taught. Based on these goals, questions (can be in the form of pictorial written exams, oral exams or tests based on demonstrations) should be prepared for pre-, mid- and post-measuring to what extent the learning targets were achieved. The PEAs can directly implement these tests as a part of the training program.

*iii. Sampling and control groups:* We recommend all participants to take the questionnaires and tests.

Even though both quantitative and qualitative type of data will be collected from all the participants, our suggestion is to analyze only the quantitative data to be analyzed for measuring overall success, while using qualitative data 1- To give detailed feedback on the effectiveness of particular trainings to the

trainers (and when necessary to Zasaka staff) 2- To be able to gain some insight in the situations where the trainings prove to be more or less effective.

*iv. Timing:* We think these measurements should be repeated each year with different groups of farmers; they should be integrated in the regular schedule of the trainings. There are two main reasons for that: 1- it is possible that the results will dramatically change depending on the characteristics and level of knowledge of the trainees and the quality of trainers. So, measuring the effectiveness of trainings only one year will not be a convincing proof of impact. 2- As one of the main purposes of this research is to improve the current practices of Zasaka, the way we structure these tests will enable Zasaka to figure out what specifically works and not works in their trainings and to intervene immediately.

#### v. Key findings from the evidence review

##### **Kirkpatrick's training evaluation model**

We were greatly inspired by Kirkpatrick's model while designing the training evaluations. Kirkpatrick's training evaluation model is the most commonly used evaluation method for measuring training effectiveness. The model breaks down the outcome evaluation into four levels: 1- Reaction, 2- Learning, 3-Behavior, 4- Results (Johnson, 2006). A slightly different version of the model, called Kirkpatrick-Phillips model also includes a fifth level: Return on investment.<sup>3</sup>

- Each level of evaluation helps to interpret the results of the next level (Johnson, 2006).
  - Level 1-Reaction:
    - Assesses reactions towards the overall program as well as its specific components such as instructors, topics, teaching methods, teaching materials.
    - Data collection methods: Questionnaires including both close-ended items including rating scales and open-ended items. Five-point rating scales with a "neutral" option are suggested. It is better to keep the questionnaires anonymous.
    - Timing: Immediately after the program.
    - Sample questions:
      - "In your view, what were the three most important weaknesses of the program?"
      - "In your view, what were the three most important strengths of the program?" (Johnson, 2006)
  - Level 2-Learning:
    - This level assesses if learning outcomes have been achieved or not. Learning outcomes can be categorized into three groups: Knowledge, attitudes and skills.

<sup>3</sup> In the research design we recommended to focus only on the first two levels of Kirkpatrick's model, because we think the best way of measuring behavioral change and change in income that can be attributed to the trainings, is to include an extra comparison group of non-Zasaka farmers who only received trainings from Zasaka and compare their behavioral change and change in income with the original control group (non-Zasaka farmers who did not use any Zasaka products or services).

- Data collection methods:
      - Knowledge: Measured by achievement tests.
      - Attitudes: Measured by questionnaires including closed and open items (very similar to Level 1).
        - Sample question: “How do you feel about diversity at the workplace?”
      - Skills: Measured by performance tests where the trainees need to demonstrate a process or create a product .
    - Timing: Immediately after the program (Johnson, 2006).
  - Level 3-Behavior:
    - Measures “the transfer of knowledge, skills and attitudes from the training context to the workplace.”
    - Data collection methods: Surveying or interviewing the trainee, her supervisor and/or subordinates, work observations, focus groups. Regarding surveys, one popular survey design that is being used is the *retrospective survey design*. In this design, the respondents are asked to reflect on their past behaviors and compare it to their current behaviors (it can be also asked to the person who observes the trainee closely). Specific examples should be demanded from the respondents. It can be done in a questionnaire or interview format.
    - Timing: Behavior changes must be assessed after some time elapsed after the training.
    - In many situations, evaluation at these three levels can be sufficient to prove usefulness of the program (Johnson, 2006).
  - Level 4-Results
    - Assesses if the training achieved its final goals such as improved quality of work, higher productivity, reduction in wasted resources, increased sales, greater job satisfaction...
    - It is the most difficult level to assess because there are often many additional causal variables that had a role in creating the results, it is difficult to isolate the role of the trainings.
    - Data collection method: Often quantitative data is collected through questionnaires or document reviews.
    - Timing: Must be assessed after some time elapsed after the training, this time may be quite long depending on the results that want to be measured. **Interrupted time-series design** is particularly suggested starting from pre-test (Johnson, 2006).

## 5. REFERENCES

- Bain & Company (2010). A passion for Social Change. Bain & Company's Social Impact Report 2009–2010. Retrieved from [http://www.bain.com/bainweb/PDFs/Bain\\_Social\\_Impact\\_Report\\_2009-2010.pdf](http://www.bain.com/bainweb/PDFs/Bain_Social_Impact_Report_2009-2010.pdf)
- Bain & Company. The Net Promoter System. Retrieved from <http://www.netpromotersystem.com/about/index.aspx>
- Clark, Andrew E. and Oswald, Andrew J. (1995). Satisfaction and comparison income. Retrieved from <http://www.andrewoswald.com/docs/jpub.pdf>
- Duflo, Ester; Kremer, Michael and Robinson, Jonathan (2008). High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya. Retrieved from <http://www.povertyactionlab.org/publication/how-high-are-rates-return-fertilizer-evidence-field-experiments-kenya>
- Duncan, Greg J. and Petersen, Eric (2000). The Long and Short of Asking Questions about Income, Wealth, and Labor Supply.
- Fields, Gary S. (1993) Poverty Changes in Developing Countries. Retrieved from <http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1483&context=articles>
- Hakimian, H. and Teshome, A. (1993). Trainers' Guide: Concepts, Principles, and Methods of Training With Special Reference to Agricultural Development. Vol. 1. FAO, Rome.
- Johnson, Burke R. (2006). The Kirkpatrick Model of Training Evaluation. Personal Collection of Burke R. Johnson's., The University of South Alabama, Mobile, AL, USA. Retrieved from <http://www.southalabama.edu/coe/bset/johnson/660lectures/Kirk1.doc>
- Meyers, Bruce D. and Sullivan, James X. (2003). Measuring the Well-Being of the Poor Using Income and Consumption. Retrieved from [http://www3.nd.edu/~jsulliv4/inc\\_cons.pdf](http://www3.nd.edu/~jsulliv4/inc_cons.pdf)
- Raab, M., & Pasqual, L. (2011). Campaigns to End Violence against Women and Girls Authors. UN Women.
- Sweeney, D. (2009). Show me the Change: A review of evaluation methods for residential sustainability behaviour change projects.
- Tesfaye, T., & Karippai, R. S. (2010). Farmers Training Effectiveness in Terms of Changes in Knowledge and Attitude: The case of Holeta, Melkassa and Debre zeit Agricultural Research Centres, Ethiopia. *Journal of Agricultural Extension and Rural Development*, 2(5), 89-96.
- USDA. (2008). Soil Quality Indicators. Retrieved from <http://doi.org/10.1016/j.jhazmat.2011.07.020>
- USDA. (n.d.) Measuring Soil Quality. Retrieved from [http://www.nrcs.usda.gov/Internet/FSE\\_DOCUMENTS/nrcs142p2\\_053155.pdf](http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_053155.pdf)
- World Bank. Measuring Living Standards: Household Consumption and Wealth Indices. Retrieved from [http://siteresources.worldbank.org/INTPAH/Resources/Publications/Quantitative-Techniques/health\\_eq\\_tn04.pdf](http://siteresources.worldbank.org/INTPAH/Resources/Publications/Quantitative-Techniques/health_eq_tn04.pdf)
- World Bank (2001). World Development Report 2000/2001 Attacking Poverty. Retrieved from [http://www3.nd.edu/~jsulliv4/inc\\_cons.pdf](http://www3.nd.edu/~jsulliv4/inc_cons.pdf)