

Feedback Matters

Evidence from Agricultural Services

Maria Jones
Florence Kondylis



WORLD BANK GROUP

Development Research Group

Impact Evaluation Team

July 2016

Abstract

Feedback tools have become ubiquitous in the service industry and social development programs alike. This study designed a field experiment to test whether eliciting feedback can empower users and increase demand for a service. The study randomly assigned different feedback tools in the context of an agricultural service to document their impact on clients' demand and shed light on the underlying mechanisms. The analysis shows large demand effects, in the current and following growing periods. It also documents large

demand effect spillovers, as other non-client farmers in the vicinity of treated groups are more likely to sign up for the service. To disentangle pure supply-side monitoring from demand-side accountability effects, additional monitoring was randomly announced to extension workers across treatment and control communities. Extension workers do not exert significantly more effort in villages where additional monitoring takes place. The study concludes that farmers' taste for "respect" leads their higher demand for the service.

This paper is a product of the Impact Evaluation Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at fkondylis@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Feedback Matters: Evidence from Agricultural Services

Maria Jones and Florence Kondylis *

Keywords: agriculture, agriculture extension, fee for service, citizen engagement

JEL Codes: O13, Q16

1 Introduction

How do feedback mechanisms affect users' demand for services? Empowering users to demand better services has become a keystone of social programs. International development organizations such as the World Bank (WB) advocate the use of participatory users' surveys to enhance service provision in their projects.¹ The idea is that users are best placed to monitor service delivery, and that their involvement in giving feedback will lead to higher empowerment, thus closing the accountability loop (Banerjee et al., 2010; Björkman and Svensson, 2009; Ringold et al, 2012). The spread of users' feedback tools is not restricted to public interest programs; customer surveys are widely used by the service industry as a way to retain customers and boost satisfaction.

Despite this enthusiasm for beneficiary monitoring, little is known as to whether establishing feedback loops can in fact affect users' demand for services. Even less explored are the channels through which feedback mechanisms increase demand for services. Typically, theories of change associated with users' feedback loops assume that satisfaction and demand for the service improve through both supply-side effects, resulting from improved accountability, and demand-side effects, as users are empowered to demand better service (World Bank, 2003). However, there is little evidence on the respective roles of pure monitoring relative to users' empowerment in improving service delivery.

* Development Economics Research Group, World Bank. Maria Jones, mjones5@worldbank.org, and Florence Kondylis, fkondylis@worldbank.org. We thank Jasmeen Batra and Cindy Sobieski for superb management of all field work, and Paula Gonzalez for excellent research assistance. We are grateful to the Ministry of Agriculture and Animal Resources in Rwanda and the staff from the Land Husbandry, Water Harvesting, and Hillside Irrigation Project for their leadership in this work. We thank One Acre Fund Rwanda for their hard work in implementing the feedback interventions and granting us unlimited access to their data. The Global Agriculture and Food Security and the i2i fund provided generous funding. We thank Jishnu Das, David Evans, Marcel Fafchamps, Xavi Gine, Steve Haggblade, Stuti Khemani, Arianna Legovini, Emilia Soldani, Daniel Stein, for their comments at various stages of the study, as well as numerous seminar audiences.

¹ This approach is often referred to, in policy circles, as Participatory Monitoring and Evaluation (PM&E), and includes Citizen Report Card and Beneficiary Assessment, as outlined here: <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALDEVELOPMENT/EXTPCENG/0,.contentMDK:20509352~menuPK:1278203~pagePK:148956~piPK:216618~theSitePK:410306.00.html>.

Our thesis in this paper is that demand-side empowerment matters for improving demand for a service, above and beyond the simple supply-side effects of feedback mechanisms. We posit that feedback tools can increase demand for a service not just by increasing monitoring, and therefore the quality of a service, but also by signaling to the users that the provider respects them. We set up a large field experiment to test our hypotheses, in which we formally document the impact of feedback tools on users' demand in the context of agricultural services for 180 farmer groups in rural Rwanda. We randomly assign two types of feedback tools to farmer groups that are subscribed to an agricultural input and extension service. Our design allows us to capture two categories of treatment effects of interest. First, we capture the combined empowerment and monitoring effects of both feedback tools. Second, we separate out pure monitoring effects from user empowerment effects by announcing to extension workers that their work is being monitored, in both treatment groups ("true" announcement) and control groups ("false" announcement).

We track the impact of these different feedback and monitoring tools on a sample of 1,460 farmers. Given the nature of our interventions, we primarily rely on administrative data to establish changes in demand as a result of the different treatments and to track demand effect spillovers to farmers who were non-users at baseline. We also use survey data to document changes in farmers' use of and satisfaction with the service, as well as empowerment to demand better service.

We find that feedback tools help sustain demand for the service among current clients and increase demand among non-users. Farmer groups offered the opportunity to provide feedback were half as likely to have members leave the service the following year as control groups. Perhaps surprisingly, this demand effect spills over to non-users in the vicinity of the treated groups, who are more likely to sign up in the following season. These effects are large and robust to multiple hypotheses testing, implying an increase in group size of up to 3

additional members in the subsequent season (0.69 SD). These effects on revealed demand for the service are corroborated by self-reported changes in demand for and attitudes toward the service. We uncover interesting heterogeneity of impact across gender. Namely, women react more positively to the feedback tools than men. Given that women had lower baseline demand for training and have lower access to land and productive assets than men, this finding is consistent with the idea that fee-based agricultural services may exclude the poorest farmers and thus may not be aligned with social objectives.

These findings are not inconsistent with a pure supply-side monitoring story, in which feedback tools improve service provision by simply signaling higher costs of shirking to extension officers (Prendergast, 1999). Randomly announcing the presence of scorecards to extension workers allows us to rule out a pure monitoring story. Extension workers do not overwhelmingly respond to our intervention by exerting more effort in villages where additional monitoring was announced. We therefore conclude that the large impacts feedback tools have on farmers' demand for agricultural services are primarily attributable to farmers' empowerment, or taste for respect.

The remainder of the paper is organized as follows. We place the experiment in the context of agricultural services in Rwanda and provide background on beneficiary feedback in Section 2. Section 3 details the experimental design and data sources. Section 4 presents the main empirical results. Section 5 concludes.

2 Motivation and Context

This section describes the general and study-specific contexts of agricultural extension services, and places our study in the context of beneficiary feedback mechanisms used in development programs.

2.1 Agricultural services

Agricultural services offer a context particularly relevant for policy. Our experiment provides evidence on the potential for simple user feedback tools to address two main pitfalls of agricultural services. First, sustaining farmers' interest in the service may be difficult beyond the first year. Evidence from successful input subsidy schemes in rural Kenya (Duflo et al 2011) shows that, despite large returns to fertilizers in the first year, fertilizer adoption collapses in the second year. Second, private extension services typically fail to secure farmers' interests in training in the long run, as they focus more on helping farmers cope with idiosyncratic shocks through just-in-time advice (Davis et al., 2014). Third, the incentives of a non-governmental service provider imply that trainings may be tailored to relatively more sophisticated farmers, excluding the more marginalized farmers (Alwang and Siegel 1994; Anderson and Feder, 2003; Axinn, 1988; Saito 1994).

The literature on agricultural extension systems emphasizes demand-side constraints to private (or fee-based)² extension provision.³ Feder et al. (2001) and Anderson and Feder (2004) provide an exhaustive survey of extension systems. The reviews highlight the tendency for general participation to drop under fee-based extension services, at the risk of excluding the poorest farmers, and that accountability mechanisms are needed to overcome potential conflicts of interests in the supply of information. For instance, fee-based extension may exclusively advertise the merits of inorganic fertilizers to boost input sales. Yet, fee-based extension services may also be more client-oriented, per their private sector logic. These concerns are particularly salient for policy as markets for agricultural services are

² In the rest of the text, we refer interchangeably to non-governmental and private fee-based extension services, since in our context the extension services are delivered against a fee by an NGO.

³ There are many institutional and political reasons governments are eager to retain centrally administered agricultural extension systems; see Axinn (1988) for a detailed description of the tradeoffs associated with different extension delivery systems.

typically thin in rural SSA. We take these observations seriously and test the extent to which stronger feedback mechanisms can be leveraged to combat dropouts in the context of a fee-based extension service in Rwanda.

2.2 Country context

Increasing agricultural productivity is central to the Government of Rwanda's Economic Development and Poverty Reduction Strategy. The agricultural sector is the sector with the greatest potential to reduce poverty in Rwanda. Approximately 85% of the population lives in rural areas and rural poverty is persistent. The rural poverty rate is 48.7%, compared to 22.1% in urban areas.⁴ Rwanda is one of the most densely populated countries in Africa, and land is scarce. In our sample, male clients live in households that farm 0.46 ha on average, female clients 0.36ha. Despite large efforts to convert non-cultivated areas into arable land, the scope to increase cultivated area is limited.

In a context where land is scarce, increasing agricultural productivity is essential for sustainable poverty reduction and economic growth. The public agricultural extension system in Rwanda has limited outreach, providing information to less than one-third of rural households (MINECOFIN, 2013). Poor quality of service delivery and informational content are also important constraints. Instead of increasing its investment in public extension services, the government chose to transition from service provider to “facilitator of the private sector”, as capacity grows and private sector investment increases.⁵ (Ministry of Agriculture and Animal Resources – MINAGRI, 2013).

⁴ These figures were published by the Ministry of Finance & Economic Planning (MINECOFIN) in 2013.

⁵ Strategic Plan for the Transformation of Agriculture in Rwanda Phase III

2.3 Agricultural service provider

One Acre Fund (OAF), known in Rwanda as Tubura, offers farmers a bundle of services, including farm inputs delivery and agricultural extension, on credit.”⁶ OAF coverage is substantial, reaching 130,400 Rwandan farm families in 2013. The agricultural extension component of the service consists of 11 training sessions over the course of the calendar year, on topics such as: planting, urea application, nurseries, crop disease identification, safe use of pesticides, composting, nutrition, harvest, and storage. Trainings are delivered at the farmer group level, with Field Officers (extension officers) providing in-field follow-up as needed.

OAF provides a fee-based service, aiming to achieve full recovery of field expenses through farmer repayment. Globally, in 2013, 73% of OAF’s expenses were covered through farmer repayment.⁷ In Rwanda, OAF clients pay an annual membership fee of approximately \$5 (RWF 2,500 in August 2012) and 19% interest on loans for agricultural inputs. On average, a farming client takes a \$48 loan from OAF in the main season: \$35 dollars for inputs, \$5 of membership, and \$8 in interest.⁸ Membership benefits include: agricultural extension trainings, delivery of inputs to the cell level,⁹ weather and yield insurance, and in-field follow-up. In the sample areas, we estimate that 3-4% of farmers were OAF members in the lead-up to the experiment. OAF Field Officers (FOs) serve as both agricultural extension agents and loan servicers, delivering trainings to farmers and collecting payments.

⁶ <http://www.oneacrefund.org/our-approach/program-model/>

⁷ Figures on repayment as a proportion of expenses are not reported at the country level. In 2013, OAF had field operations in Rwanda, Burundi, Kenya and Tanzania.

⁸ Interest rates are not waived even if a client can pay the amount upfront, to cover operating costs.

⁹ Rwanda’s administrative divisions, from largest to smallest: districts, sectors, cells, and villages (imidugudu).

2.4 Feedback mechanisms in development programs

The intervention we study is akin to a ‘citizen report card’ experiment. This concept was first developed by civil society organizations in Bangalore, India, in 1993 in response to concerns about the quality of public services. Citizen report cards provide citizen feedback on various programs, and are meant to serve as a social accountability mechanism that identifies poor performance and stimulate improvement (World Bank, 2001). They are part of an approach to making services work for the poor that emphasizes “client power.” This approach has gained prominence in the last decade; the World Development Report 2004 “Making Services Work for Poor People” emphasizes the need to increase poor citizens’ voice and participation.¹⁰ However, as Duflo and Banerjee (2006) document, despite widespread enthusiasm, few programs that increase beneficiary control have been rigorously evaluated, and the evidence so far is mixed. In addition, feedback mechanisms appear to yield varying degrees of impact on quality and quantity of service provision, as well as on usage. To some extent, this could be attributable to differences in implementation protocols, which are typically not formally tested. For instance, as Duflo and Banerjee (2006) discuss, mediated vs. mechanized protocols could mean the difference between success and failure for this class of interventions.¹¹ While, ultimately, all citizen report cards and feedback mechanisms prompt users to give feedback on services, there is no consensus on the best way to elicit that feedback.

Most citizen report card interventions in development programs target the delivery of health and education services. The few rigorous evaluations that have been done are also specific to those sectors. Bjorkman and Svensson (2009) show that community monitoring of

¹⁰ World Bank Development Report (2004).

¹¹ As cited in Banerjee and Duflo (2006), see examples of mediated feedback process in Kremer and Chen (2001) and mechanized process in Duflo et al. (2011).

public health facilities—through a citizen report card implemented in a random subset of communities in Uganda—improves the quality and quantity of health care, as health unit staff exert greater effort to serve the needs of the community. There are few examples of citizen report cards in agriculture. One exception is Mogues et al (2009), who find high rates of satisfaction with extension services in Ethiopia: 92% of men and 94% of women who received extension or expert advice report being ‘very satisfied’. However, only 8% of those farmers reported adopting any new practice or use of inputs (whether improved or conventional) over the last two years. Interestingly, looking at the descriptive statistics from our scorecard experiment, we find similar levels of satisfaction with the fee-based extension service, with under 3 percent of farmers reporting unhappiness with any specific training.¹² A central limitation of the methodology is that farmers who do not attend extension trainings do not have a voice. To circumvent this issue, we design an experiment in which the intervention of interest is giving access to a feedback tool, and we do not attempt to analyze satisfaction levels from the data collected in the process.¹³ Instead, we are interested in the impact giving farmers a voice has on their demand for extension services.

The literature on citizen empowerment presents mixed evidence on the impact of feedback loops or scorecards in boosting citizens’ empowerment. At best, the evidence suggests that these mechanisms offer a partial solution to improving the quality of service delivery. In our context, the main concern for both the service provider and the government in ensuring quality of service delivery is that FOs’ (extension workers) incentives may not be aligned with farmers’ interests. Indeed, repayment rates on loans are the main metric of an FO’s

¹² The response rate was high, ranging from 56 to 97 percent across the different formats of scorecards. Satisfaction rates do not appear sensitive to response rate. Satisfaction levels are computed conditioning on a farmer attending a specific training.

¹³ In addition to the absence of a counterfactual in our pure control group, we compare different feedback mechanisms side by side. As a result, we could not devise a common metric of satisfaction across treatment arms, making the exercise fruitless. OAF used the feedback for accountability purposes; to identify problems and instances of fraud, rather than as a general satisfaction measure.

performance evaluation, as well as her total outreach (number of clients); this raises concern over the content of their interactions with farmers. Introducing “softer” measures of performance in their assessment is therefore of interest both in reinforcing satisfaction with and demand for the service.

3 Experimental Design and Data

The experiment took place in 3 of Rwanda’s 30 districts: Karongi and Rutsiro districts in the Western province and Rwamagana district in the Eastern province. The sampling frame is all registered clients of OAF for Season 2013A, comprising 228 farmer groups.¹⁴ We randomly sample 180 of the 228 groups, and 8 farmers within each of those groups.¹⁵ With 1.8% attrition across two survey rounds, the final sample for analysis is 1,460 farming clients, in 20 cells.

We worked with OAF and the Ministry of Agriculture to introduce and test a variety of farmer feedback tools between September 2012 and June 2013, thus covering two agricultural seasons.¹⁶ We devise a randomized-controlled trial (RCT) to test two types of feedback tools: scorecards and logbooks. A third intervention cross-cuts the scorecard treatment and explores the impact of announcing scorecards to FOs to capture the pure monitoring effect of adding a feedback tool. Table 1 summarizes the study design.

¹⁴ While the sampling frame is the population of OAF clients in 2013A, we use the OAF client registry in the subsequent two seasons (Season 2013B and Season 2014A) to track retention of members over seasons and the addition of new members.

¹⁵ The average of 8 individuals per farmer group gives us sufficient power to detect fairly small effects; at 80% power and assuming intracluster correlation of 0.1, the minimal detectable effect size is .43 standard deviations on the outcomes of interest for as few as 12 control groups (this is the most conservative estimate of power for the experiment, as comparisons between treatment arms will be higher powered than comparisons to the pure control group).

¹⁶ Cell is an administrative division in Rwanda above village, and the level at which OAF organizes clients.

The scorecard is a structured feedback tool, designed to anchor farmers' expectations of the content and quality of the service through a series of topic-based questions with multiple choice answers. Scorecards are supply-driven and administered by our survey team. In contrast, the logbook records farmers' experience on similar issues in an open-ended manner, without providing information on expected standards or service, nor guiding answers. The logbooks are demand-driven; farmers could log their comments anytime.

Within our study sample, we randomly assign 96 groups to the scorecard treatment, while the remaining 84 serve as control. Scorecards were administered every three months over the study period; each scorecard surveyed farmers' experience with three different types of trainings that OAF FOs were mandated to hold over the last three months.¹⁷ The scorecards were then used by OAF to assess the quality of their service, in a way that anonymized the group and farmers who provided data but not the responsible FO. The results (aggregated at the group level) were shared with each treated group in the subsequent round of scorecard data collection, just before each interview.

We next overlay the logbook treatment to scorecard treatment: 132 randomly chosen groups receive the logbook, of which 72 also receive the scorecard intervention and 60 do not, while 48 serve as control, of which 24 also receive the scorecard.¹⁸

Thirdly, we overlay an "announcement" intervention to our scorecard treatment arm. OAF informed its FOs that scorecards would be implemented at random intervals during the next two agricultural seasons. They made the announcement in half of the groups assigned to the

¹⁷ We randomly assigned farmers to three different types of scorecards: visual scorecards administered in paper format during a group meeting; visual scorecard administered on a tablet during home visit; and a phone scorecard, where each farmer was interviewed on the phone. We cannot reject that all formats had similar effects (not reported).

¹⁸ As in the scorecard treatment, we randomized three different types of logbook collection methods: by the monitoring and evaluation specialist, by the local field manager, or through a hotline free call-in service. We cannot reject that all formats yield point estimates that are within confidence interval of each other (not reported).

scorecard treatment (“true” announcement), and half of the groups assigned to control (“false” announcement). This provides a “placebo” where the FOs are told that scorecards might happen but in fact do not. Comparing announcement to no announcement within the scorecard treatment provides a test of the extent to which awareness of the scorecards can itself boost efforts of the FO in training these specific farmers groups, while additionally comparing true and false announcement tells us about FOs’ ability to observe the scorecards. While we control for logbook treatment assignment dummies in our analysis of the announcement effect, we are not powered to study their interactions.

3.1 Data

We use two sources of data: administrative data from our partner NGO, OAF, and survey data administered on a random subset of our study sample. First, we exploit OAF’s administrative client database for two years, 2013 and 2014. The client database includes the names of all farmers who register for OAF membership in the season, the name of the group to which the farmer belongs, the OAF Field Officer and Field Manager responsible for the group, the total amount of credit requested by the client, the quantity requested of specific agricultural inputs,¹⁹ the number of repayments, and the proportion of the loan repaid. We use the 2013 client database as our sample frame. We track whether these farmers continue with the service by merging with the 2014 client database. If a client in 2013 does not register for the service in 2014, she is considered a dropout. If a farmer was not on the client list in 2013A, but registers in 2014, we consider her to be a new member. Some clients registered in 2013 (in and outside our study sample) switch farmer group in 2014; we do not consider these farmers to be either dropouts or new members. Hence, we report changes in membership net of these switches. For instance, if two clients were registered in group X in

¹⁹ OAF offers DAP, NPK-17, NPK-22, and Urea.

2013 but switch to group Y in 2014, we count them as members of group X in 2014 regardless of group X and Y's treatment status. Our computation of total group size in 2014 is therefore not the equivalent of computing (size in the previous year + new members – dropouts), but instead (size in the previous year + new members + switchouts – dropouts – switchins), where switchouts are clients who left the group for another one, and switchins join the group from another one. In our group-level analysis, we follow the initial 2013 group composition and apply these definitions of dropouts and new members.

Second, we collect a multi-module household survey among OAF clients in our study sample in May–August 2013.²⁰ The instrument was designed to generate quantitative data on agricultural production, participation in agricultural extension services, perceptions of agricultural extension providers, and socioeconomic characteristics over the course of two visits.

3.2 Farmer characteristics

OAF clients in the sample areas took loans averaging \$56 (RWF 38,641) in 2013. Female clients take significantly smaller loans, \$48 on average for 2013, compared to \$60 for men.

19 OAF FOs serve our study area. One FO is responsible for 12 farmer groups on average (minimum 2, maximum 18), a total of 134 clients (minimum 20, maximum 209), and are overseen by 5 FMs. FMs manage an average of 6 of the sampled FOs each (minimum 2, maximum 9).²¹

Of the sampled households, 73% are male-headed. Household heads are on average 45 years old, and most (77%) have less than a primary school education. The household head is the OAF client in 86% of households. Compared to male OAF clients, female clients are

²⁰ Fieldwork was implemented by an independent survey firm.

²¹ Statistics come from OAF administrative records and refers to all OAF clients in LWH sites.

members of households that have less farmland, and have a less-educated and older household head.

3.3 Balance

We verify that observed socioeconomic characteristics are balanced across our treatment and control groups. For brevity, we only report results for the main assignment variables: we pool together the two feedback treatment groups (scorecards and logbooks) in Table 2, and announcement in Table 3, although the results hold for all treatment arms. We find that most individual and group-level characteristics are balanced across treatment and control groups, with the exception of the gender of the head and the age of the head in our combined treatment group (Table 2), and construction material of household walls in the announcement treatment (Table 3). The differences are small, and we control for these characteristics in our regression analysis.

3.4 Stable Unit Treatment Variations Assumption (SUTVA)

Ideally, we would have assigned our intervention at the FO level. However, given OAF's administrative structure and the resulting low number of FOs in our sample area (19), we could not satisfy sample size requirements to statistically detect a reasonable effect size. As FOs serve both treatment and control groups, we worry that our intervention may displace effort from one farmer group to another and, thus, that our study protocol may violate SUTVA. A mitigating factor is that FOs are incentivized on the basis of farmers' repayment rate overall. The feedback tools were introduced as an additional check on farmer satisfaction with the service, but not explicitly tied to individual FO performance indicators. Hence, there is little risk that they would adjust on the extensive margin and choose to altogether neglect some groups to the benefit of others, and risk lowering their overall repayment rates. In fact, we find that farmers' repayment rates are not affected by

our intervention, which suggest that FOs did not respond by not servicing control farmers groups (col. 1, Annex Table 1). As an additional check, we verify that training attendance, aggregated up to the group level, is not affected by our interventions. Aggregating attendance at the group level gives us a measure of access to training within a given group. For this, we aggregate individually reported attendance in training at the group level, taking the maximum value of the dummy for attendance and the average number of trainings attended by group members. The results (Cols. 2 and 3, Annex Table 1) suggest that trainings were equally offered to all groups regardless of their assignment to our interventions.

We expect that FOs would rather adjust their effort on the intensive margin, by providing better service in groups where farmers demand it, while still attending to other groups at pre-intervention levels. However, different FOs may respond differently to the treatments, and, by nature of the randomization, different FOs served different proportions of treatment and control groups. We control for FO heterogeneity in all our regression analysis. We will therefore estimate within-FO variations in farmers' demand for the service as a result of our interventions.

4 Results

4.1 Farmers' demand for extension services

We first test whether our intervention affects farmers' decision to sign up for the service in the following year. This is the most objective measure of farmers' demand we have access to, and it is sourced from administrative data from our partner OAF in Rwanda. In addition to tracking take up decisions among existing clients, we measure the impact of the intervention on locally-placed non-users. As described in section 3.1, we measure the incidence of drop outs and new member enrollment at the group level. Our empirical strategy is based on random assignment of farmers groups into the different treatments—Logbooks (L), Scorecards (S), Announcement (A), and their interactions (LS , AS). We estimate the following equation:

$$y_{jkl} = \alpha + \beta_{FL}L_j + \beta_{SL}S_j + \beta_{FSL}LS_j + \delta_{AL}AL_j + \delta_{ASL}AS_j + \theta_k + \epsilon_i \quad (1)$$

where, j indexes the farmer groups, k indexes the FO for group j , and l indexes the outcome. θ_k represents FO effects (for which FO dummies are included), and ϵ_i is an individual random error term. Results are shown in Table 4, where the top panel reports the total effect of being in each possible treatment arm, and the lower panel reports a number of additional statistics, including the p-values of differences in relevant treatment effects to guide inference.

First, we check whether our intervention did affect the structure of farmers' groups. Groups can merge and divide up, leading to attrition from one season to another; our intervention may have affected this dynamic. We estimate (1) on a dummy indicating group survival (takes value 1 if the group remains active in the following season, 0 if it dies; col. 1, Table 4). The results indicate that our intervention did not affect a group's probability to survive in the following season, allowing us to interpret changes in the incidence and

prevalence of dropouts and new enrollments. We additionally check for balance in group size in the season contemporaneous with the launch of our intervention, Season A 2013 (col. 2, Table 4). Reassuringly, we find that group size was not significantly different across treatments and control. These null results are confirmed by tests of joint significance on all treatment variables in both equations (cols 1 and 2, Table 4).

Scorecards and logbooks have similar effects on group-level incidence of dropouts. Groups offered feedback mechanisms were 29-44 percentage points less likely to have at least one member drop out, deducted from an 88 percent incidence in the control group (col. 4, Table 4). The intensive margin indicates a large reduction in dropouts as a result of our intervention as well, with groups offered logbooks losing 1.94 fewer members than the control (0.48 standard deviation), a 46 percent decrease, (col.5, Table 4). Scorecards led groups to lose 1.03 fewer users, although the effect is not precisely estimated (0.25 of a standard deviation). The finding that feedback tools reduce dropouts at the extensive margin is robust to multiple-hypothesis testing, as indicated by a test of joint significance on all treatment variables. Our results do not point to complementarities between logbook and scorecard, as overlaying these feedback tools does not affect the results.

To further understand the way feedback tools affected the extensive and intensive margins of dropouts, we turn to the distribution of number of dropouts across treatments and control (Figures 1, 2). We see that providing feedback tools shifted the distribution of dropouts towards zero and reduced the incidence of higher dropout rates.²² This unambiguously reduced the number of dropouts in the logbook treatment (Table 4, col. 5). Interestingly, we see that groups that were offered a feedback mechanism were more likely

²² Kolgorov-Smirnov tests of the equality of distributions show that the distribution of dropouts is significantly different under the scorecard treatment, compared to the pure control (exact p-value 0.08). The distribution for the logbook treatment, compared to the pure control, is marginally insignificant (exact p-value 0.13).

to either have no dropouts at all, or have a lot of dropouts, relative to the status quo. These results suggest that, in the presence of a feedback mechanism, farmers' decision to stay or leave the service was more coordinated across members.

Demand effect spillovers.

We next test whether providing users with feedback tools had any impact on non-users in the vicinity. The underlying hypothesis is that the lack of opportunity to influence the delivery of OAF's agricultural services presents a significant barrier to farmers' take up.

Again, we are interested in capturing both the extensive and intensive margins of expanded membership across treatment and control. Using the administrative data, we regress both a dummy for having new members join the group and the number of number of new members on our treatment assignment variables, as described in equation (1). We find evidence of large, positive spillover effects to non-users in the vicinity of treated groups (cols 6 and 7, Table 4). Groups that were offered any channel of feedback are 26-28 percentage points more likely to attract new members (col. 6, Table 4); these are very large effects relative to an 8 percent chance in the control group. Looking at the intensive margin, we find logbooks have a larger, more precise effect than scorecards. Logbooks add on average 1.12 new members to a group (significant at the 1% level), or 1.35 of a standard deviation, relative to 0.61 for scorecards (not significant). However, we cannot statistically tell these coefficients apart in this single equation.²³

Taken together, these results show that providing a simple feedback tool such as a logbook positively affected users' sustained demand for the service and had large, positive demand spillovers for non-users. This indicates that our intervention addressed a significant barrier to adoption of this service—a lack of accountability. Neighboring client farmers

²³ A Wald test of joint significance indicates that most of these results are robust to multiple hypothesis testing.

gaining access to channels of feedback may have signaled to non-users that OAF will take their opinion of the service seriously. The suggestive evidence that logbooks were more effective than scorecards may have to do with the fact that groups physically kept the logbook in the community, while scorecard feedback was collected by others.

4.2 Channels

To shed light on the channels underlying the large effects of feedback tools on farmers' revealed demand for agricultural services, we (1) track the impact of our intervention on self-reported measures of demand, and (2) add an experimental arm to our study to test a pure monitoring (supply-side) effect against the user empowerment channel.

In most specifications, we estimate the following model on men and women separately:

$$y_{ijk} = \alpha + \beta_F L_j + \beta_{SI} S_j + \beta_{FSI} LS_j + \delta_A A\bar{S}_j + \delta_{AS} AS_j + \theta_k + \epsilon_i \quad (2)$$

where, i indexes the user, j indexes the farmers group, and k indexes the FO serving group j . θ_k represents FO fixed effects, and ϵ_i is an individual random error term. To draw statistical inference, standard errors are clustered at the farmer group level, using White's correction, to account for the unit of randomization and allow for correlated residuals within farmers group. We estimate (2) on a range of outcomes variables along the causal path, and report the total effects of being in each possible treatment arm.

Farmers' use and perception of the service

We first look at the impact of providing a feedback tool on farmer's self-reported interactions with the agricultural service. Since male and female clients have different baseline levels of interaction, we split our sample along gender lines.²⁴ We find that providing

²⁴ We divide the sample using the gender of the OAF client. This is the same as the gender of the household head for 85.8% of households. We control for the gender of household head in the regressions.

farmers with feedback tools increased participation in OAF's extension services, with some heterogeneous effects by gender (Table 5). Men's interactions with extension services were affected at the intensive margin of their attendance in training (col. 3, Table 5). These are large effects: male clients participated in 0.67-1.23 additional trainings relative to a mean of 1.85 and a standard deviation of 2.2 in the control. We do not find any complementarities when both scorecards and logbooks are distributed.²⁵

For women, access to feedback tools affects both the extensive and intensive margins of interactions with extension services: women are 14-25 percentage points more likely to attend at least one training a year relative to 18 percent of the control (col. 7, Table 5), and attend .86 to .97 additional trainings per year (mean of 1.47 trainings attended by female clients in control). For both men and women, the impacts of scorecards and logbook are all within confidence intervals.

Next, we check whether this increase in attendance in trainings affected farmers' knowledge of the various techniques through a simple test score (Table A1). We fail to detect any impact on knowledge. This is not altogether surprising, given that the trainings promoted best practices and correct usage rather than focusing on new technologies, and that error in measuring farmers' knowledge likely biases our estimates downwards (Laajaj and Macours, 2015). We also estimate equation (1) on yields and find no meaningful impact (not reported).

Finally, we find that having access to feedback tools did not significantly affect farmers' propensity to experience and report problems with the service (Table 5, Cols. 4-5, 9-10). This is consistent with the fact that repayment rates are identical across treatment and control

²⁵ We also report these results by training and find similar results (Table A3.A and B).

(Table A2), and lends support to our identifying assumption: agents did not reallocate their effort from control to treatment communities.

Did our intervention impact farmers' opinion of the service? To shed light on this question, we turn to farmers' perception of their FO. We asked farmers a series of question about their agent's knowledge of agriculture, teaching skills, and honesty, using a {1,2,3,4} scale, 1 being very bad and 4 being very good. We pool these measures into a single index and to simplify interpretation normalize it to a binary variable, with mean 0.43 and 0.35 for men and women respectively in the control group, and standard deviations of 0.48 and 0.46, respectively. We estimate equation (1) on this index for men and women separately (Table 6). The results suggest that male clients' opinions of their extension agents did not change with access to either feedback tools. In contrast, female clients' perceptions of their extension agent improved across the board, with effect sizes of about 0.5 SD, and no significant complementarities across feedback tools.

Taken together, these results confirm that the notion that improving accountability of the provider increased farmers' interaction with the service. We also find significant gender heterogeneity, suggesting accountability was a relatively more salient concern for female clients.

Monitoring vs. empowerment effects

So far, we have shown that increasing accountability positively affects farmers' demand for an agricultural service. However, our results are consistent with both a pure (supply-side) monitoring effect and a users' empowerment story. Feedback tools could have simply signaled higher costs of shirking to FOs, who then responded by improving their service and affecting demand. In this story, whether farmers provide feedback or not relevant, it is the increased monitoring that matters.

We devise a simple test to tell apart these competing hypotheses. We randomly announce additional monitoring to extension providers in communities assigned both to receive and not to receive scorecards.²⁶ Should these effects result from a pure supply-side story, we would find the effects of only announcing to be indistinguishable from the effects of actually implementing scorecards and announcing monitoring. This would allow us to rule out the demand-side story—that feedback tools empowered farmers to demand better services.

To enhance the credibility of our announcement intervention, we restrict our test to the scorecard treatment arm, as logbooks stayed with the communities throughout the experiment and were plausibly visible to the extension agents. We cannot rule out that extension agents had full knowledge of the location of the scorecards, which would weaken our identification of a pure response to monitoring. However, the data collection was done by an independent survey team that conducted multiple surveys for other purposes in the study districts over the course of the year, so the scorecards may not have caught attention. Should extension officers be able to perfectly verify scorecard implementation, then they should not react to false announcement (i.e., announcement of scorecards in groups which actually did not receive them), and only respond to the announcement in actual scorecard groups.

There are two important limitations in interpreting the results from this test. First, it will provide results only if either the presence or announcement of scorecards has an impact on farmers' outcomes; if not, either announcing or offering feedback will have zero impact and our test will be tautological ($Y_{T_S|A} = Y_{C_S|NA} = 0$), where $(T_S|A)$ denotes being assigned to scorecards and announcement, and $(C_S|A)$ denotes being assigned to announcement but not scorecards. This is not a concern as we find that scorecards affect farmers' outcomes, either through supply or demand channels. Second, we are not sufficiently powered to precisely

²⁶ We could only run this test in the scorecard part of the experiment, as logbooks, which stay in the communities, are fully visible to the extension officers.

estimate and tell apart small differences in estimated parameters. As a result, we are left to interpret non-statistically meaningful differences. We control for the assignment variable to this “Announcement” intervention and its interaction with the scorecards treatment in all previous specifications, and results are reported in all Tables of results (Tables 4-6).

First, we find that FOs can imperfectly tell where the scorecards take place; they only sometimes adjust their performance to the announcement (i.e. we reject that $[Y_{Ts|A}-Y_{Ts|NA}]-[Y_{Cs|A}-Y_{Cs|NA}]=0$ with $|Y_{Ts|A}-Y_{Ts|NA}| > 0$, $|Y_{Cs|A}-Y_{Cs|NA}| > 0$). This is particularly clear when looking at results on male attendance, both on aggregate and by training (Tables 5 and A2.A). Should FOs have been able to *perfectly tell* where the scorecards happened, we should see that $[Y_{Ts|A}-Y_{Ts|NA}]=[Y_{Cs|A}-Y_{Cs|NA}]=0$. Our results show that this is not the case; our false announcement does elicit a positive demand response from male client farmers, who respond by attending significantly more trainings (Tables 5 and A3.A). Yet, in most specifications, these effects are too small to be statistically different from zero, and much smaller than the effect of the feedback tools themselves (although we cannot reject parameter equality).

Overall, while we find some weak evidence that announcing the scorecards affected farmers’ demand for the service both in actual treatment and control communities, these effects are much smaller than the effect of the scorecards, and mostly statistically indistinguishable from zero. FOs could not perfectly “game” the presence of scorecards, which lends support to the idea that FOs did not overwhelmingly respond to our intervention by exerting more effort in villages where additional monitoring took place. As the effects of announcing monitoring are small relative to the effects of scorecards alone, we conclude that the large impacts feedback tools have on farmers’ demand for extension services are primarily attributable to farmers’ empowerment and taste for respect, not to a pure monitoring story.

5. Discussion and Sustainability

We designed and evaluated a large randomized controlled trial to shed light on whether farmers' ability to provide feedback on a service received from a non-governmental service provider affects their demand. Our results show that offering feedback channels is a meaningful way to combat dropouts and increase attendance in extension trainings, and that these positive demand effects spill over to non-users in the vicinity. Farmers groups who have access to feedback tools are 28 percentage points more likely to attract new members, relative to control farmers groups that have an 8 percent chance to attract new members. This shows that feedback tools retain farmers but also address constraints to adoption among non-users.

Men and women respond differently to the feedback tools, implying that accountability is a more salient constraint for women. Offering feedback appears to be particularly effective in getting women to start interacting with the FO and improve her perceptions of the FO, while for men, it only intensified participation with no effect at the extensive margin. This may result from a combination of demand-side constraints, as women perceive the benefits of training to be low or face higher opportunity cost, and supply-side inefficiencies, as providers have little information as to what training more marginal farmers require and, therefore, face higher acquisition costs in reaching this subpopulation.²⁷

Anchoring expectations with a content-based scorecard did not have a distinguishable additional positive impact on farmers' demand. Indeed, logbooks unambiguously decreased the incidence and numbers of dropouts in the following season, while scorecards led to a lower incidence with no effect on the number of dropouts. This lends support to the notion that (salience of) information was not the mechanism through which feedback tools affected farmers' demand for the service.

²⁷ Axinn (1988) cites the need for a two-way communication channel between extension and farmers to reduce these inefficiencies.

Overall, our findings suggest that providing channels for users to voice their opinion on the service they receive can trigger higher levels of demand, by signaling respect to their clients. Implicitly, organizations that choose to establish these channels of communication are likely interested in using the feedback to guide their operations. This is important, as we find that our more intensive mode of feedback collection, the scorecard, may in fact have a negative impact on farmers' attitudes in the face of future problems with the service (not reported). Hence, feedback mechanisms need to be taken seriously by the providers, or risk being counter-productive. Hence their impact may not be warranted when imposed in a top-down fashion, such as is the case when central governments impose accountability mechanisms at the sub-national level, where services are managed.

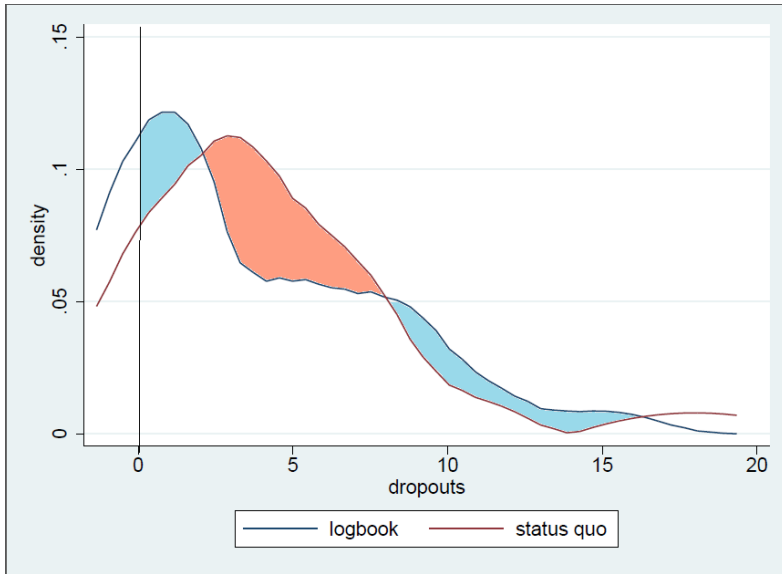
References

- Akroyd, S., and L. Smith (2007). "Review of Public Spending to Agriculture", Oxford Policy Management, July 2007.
- Anderson J.R., and G. Feder (2003). "Rural extension services." World Bank Policy Research Working Paper 2976.
- Axinn G. (1988), *Guide on Alternative Extension Approaches*, FAO, Rome.
- Banerjee, Abhijit, and Esther Duflo. 2006. "Addressing Absence." *Journal of Economic Perspectives*, 20(1): 117-132.
- Banerjee A., R. Banerji, E. Duflo, R. Glennerster, and S. Khemani (2010). "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," *American Economic Journal: Economic Policy*, American Economic Association, vol. 2(1), pages 1-30, February.
- Banerjee A., A. Deaton, and E. Duflo (2013). "Wealth, Health, and Health Service Delivery in Rural Rajasthan," *American Economic Review*, 94, 326–330.
- Birkhaeuser D., E. Robert, and G. Feder (1991). "The Economic Impact of Agricultural Extension: a Review." *Economic Development and Cultural Change*, 39(3): 607–650.
- Björkman M. and J. Svensson (2009). "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda," *The Quarterly Journal of Economics*, MIT Press, vol. 124(2), pages 735-769, May.
- Bjorkman, Martina and de Walque, Damien and Svensson, Jakob, *Information is Power: Experimental Evidence on the Long-Run Impact of Community Based Monitoring* (August 1, 2014). World Bank Policy Research Working Paper No. 7015.
- Davis, K., S.C. Babu, and S. Blom (2014). "The role of extension and advisory services in building resilience of smallholder farmers", *Building Resilience for Food and Nutrition Security: Conference Brief 13*.
- Dercon, Stefan, and Daniel O. Gilligan, John Hoddinott and Tassew Woldehanna (2009). "[The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages](#)," *American Journal of Agricultural Economics*, Agricultural and Applied Economics Association, vol. 91(4), pages 1007-1021.
- Duflo, E., M. Kremer and J. Robinson (2011). "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *American Economic Review*, vol. 101(6), pp. 2350-90.
- Evenson, R. (2001). "Economic impacts of agricultural research and extension", [Handbook of Agricultural Economics, Volume 1, Part A](#), Pages 573–628.
- Feder, G., A. Willett and W. Zijp (2001). "Agricultural extension: Generic challenges and the ingredients for solutions" in *Knowledge Generation and Technical Change: Institutional Innovation in Agriculture*. S. Wolf and D. Zilberman (Eds). Boston, Kluwer.

- Gollin, D., M. Morris, and D. Byerlee (2005). "Technology adoption in intensive post-Green Revolution systems." *American Journal of Agricultural Economics* 87(5): 1310-1316.
- Hanna, R., S. Mullainathan and J. Schwartzstein, 2012. "Learning Through Noticing: Theory and Experimental Evidence in Farming," NBER Working Paper 18401.
- Jack, K.. "Market inefficiencies and the adoption of agricultural technologies in developing countries." ATAI White Paper, CEGA (Berkeley) and J-PAL-MIT (2011).
- Kondylis, F., V. Mueller, and S. Zhu (2014a). "Seeing is believing? Evidence from an extension network experiment". World Bank Policy Research Paper 7000 (August 2014).
- Kondylis, F., Mueller, V., Sheriff, G., & Zhu, S. (2016). Do Female Instructors Reduce Gender Bias in Diffusion of Sustainable Land Management Techniques? Experimental Evidence From Mozambique. *World Development*, 78, 436-449.
- Laajaj, R., Macours, K., (2015). The reliability and validity of skills measurement in rural household surveys. Mimeo, Paris School of Economics.
- Miller, M., M. Mariola, and D.O. Hansen (2008). "EARTH to farmers: Ecological management and sustainable development in the humid tropics of Costa Rica". *Ecological Engineering*, Volume 34(4): 349-357.
- Ministry of Agriculture and Animal Resources Rwanda – MINAGRI (2013). "Strategic Plan for the Transformation of Agriculture in Rwanda Phase III. Republic of Rwanda.
- Ministry of Finance and Economic Planning Rwanda – MINECOFIN (2013). "Economic Development and Poverty Reduction Strategy 2013 – 2018: Shaping our Development." IMF Country Report No. 13/360
- Mogues, T, M.J. Cohen, R. Birner, M. Lemma, J. Randriamamonjy, F. Tadesse and Z. Paulos (2009). "Agricultural Extension in Ethiopia through a Gender and Governance Lens". IFPRI discussion paper Discussion Paper No. ESSP2 007.
- Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature* 37 (1), 7–63.
- Ringold, Dena, Alaka Holla, Margaret Koziol, and Santhosh Srinivasan (2012). *Citizens and Service Delivery*. Washington, DC: World Bank.
- World Bank (2001). "Case Study 1: Bangalore, India, Participatory Approaches in Budgeting and Public Expenditure Management." Participation Thematic Group, Social Development Department, Washington, D.C.
- World Bank (2003). *World Development Report 2004: Making Services Work for Poor People*. New York, NY: Oxford University Press.

Figures

Figure 1: Distribution of dropouts, Logbook treatment vs. Control (status quo)

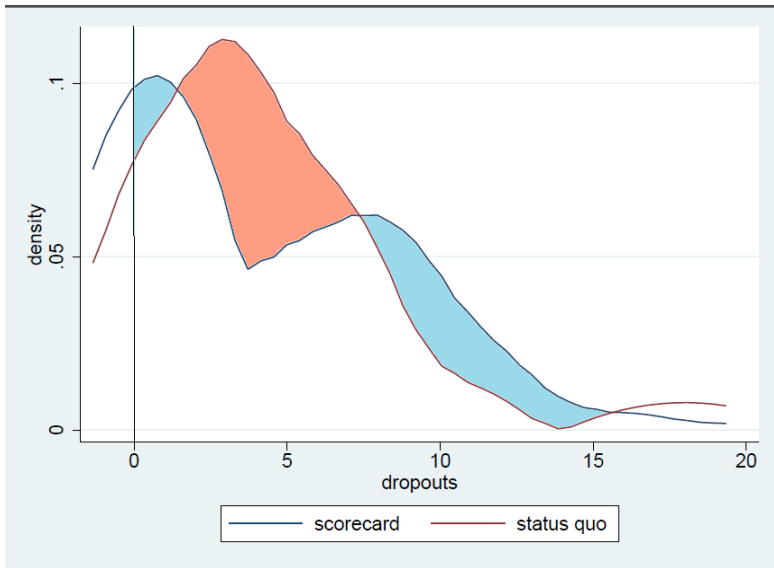


Logbook Treatment vs Status quo.

	D	P-value	Exact P-value
Status Quo	0.197	0.426	
Logbook Treatment	-0.3409	0.078	
Combined K-S:	0.3409	0.155	0.127

Note: 132 logbook treatment groups. Status quo is the 12 pure control groups.

Figure 2: Distribution of dropouts, Scorecard treatment vs. Control (status quo)



Scorecards Treatment vs Status quo.

Smaller group	D	P-value	Exact P-value
Status Quo	0.2604	0.235	
Feedback Treatment	-0.375	0.05	
Combined K-S:	0.375	0.1	0.08

Note: 96 scorecard treatment groups. Status quo is the 12 pure control groups.

Tables

Table 1: Feedback Treatment Assignment (group level)

		Logbooks			Announcement (x scorecards)	
<i>Assignment</i>		Treatment	Control	Total	Treatment	Control
Scorecards	Treatment	72	24	96	True (48)	48
	Control	60	24	84	False (42)	42
	Total	132	48	180	90	90

Table 2: Balance tests on the assignment to feedback treatment

	Treatment		Control		Difference	P_Value
	#	Mean	#	Mean		
Age of household head (years)	722	45.54	738	43.74	1.803 *	0.057
Male household head (dummy)	722	0.70	738	0.75	-.05 **	0.037
Male respondent (dummy)	722	0.59	738	0.62	-0.026	0.303
HH completed primary school (dummy)	722	0.21	738	0.24	-0.028	0.257
Number of dependents	722	1.73	738	1.68	0.047	0.594
Total land own	721	0.43	733	0.41	0.023	0.627
Main construction material	722	0.12	738	0.12	-0.002	0.944
Primary source of drinking water	722	0.27	738	0.25	0.015	0.7
Total Credit Season A 2013 (RWF)	722	33706.34	738	32106.72	1599.623	0.454
Number of clients per group Season A 2013	156	9.85	24	9.21	0.64	0.48

Critical values and standard errors clustered at the group level.

Table 3: Balance tests on the assignment to announcement treatment

	Treatment		Control		Difference	P_Value
	#	Mean	#	Mean		
Age of household head (years)	1278	44.84	182	43.13	1.712	0.229
Male household head (dummy)	1278	0.73	182	0.70	0.028	0.449
Male respondent (dummy)	1278	0.60	182	0.64	-0.04	0.298
HH head completed primary school (dummy)	1278	0.22	182	0.25	-0.032	0.387
Number of dependents	1278	1.72	182	1.59	0.127	0.334
Total land own	1274	0.41	180	0.48	-0.069	0.328
Dwelling walls are unimproved	1278	0.11	182	0.18	-.067 **	0.039
Primary source of drinking water is unimproved	1278	0.26	182	0.27	-0.019	0.747
Total Credit Season A 2013 (RWF)	1278	32273.53	182	37281.15	-5007.623	0.111
Clients per group Season A 2013	90	9.97	90	9.57	0.4	0.517

Critical values and standard errors clustered at the group level

Table 4: Client retention and demand effect spillovers (farmer group level)

	Group Dynamics			Farmers Demand			
	Group Survival	Size of group in Season A 2013	Size of group in Season A 2014	New users joined the group (Dummy)	New users joined the group (Number)	Users dropped out (Dummy)	Users dropped out (Number)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Scorecard only	0.15 [0.13]	0.14 [1.07]	1.78 [1.39]	0.28** [0.13]	0.61 [0.56]	-0.44*** [0.14]	-1.03 [1.26]
Logbook only	0.14 [0.10]	-0.04 [0.88]	3.02*** [0.90]	0.26*** [0.10]	1.12*** [0.37]	-0.29*** [0.11]	-1.94** [0.94]
False announcement	0.04 [0.08]	0.16 [0.84]	-0.26 [0.85]	0.02 [0.10]	-0.05 [0.42]	-0.02 [0.10]	0.37 [0.76]
Scorecard and Logbook	0.11 [0.11]	0.55 [0.98]	1.62 [1.04]	0.25** [0.11]	0.63 [0.40]	-0.24** [0.11]	-0.44 [1.01]
Scorecard and announcement	0.15 [0.13]	-1 [1.17]	2.13 [1.37]	0.33** [0.13]	1.34** [0.55]	-0.54*** [0.14]	-1.8 [1.30]
Observations	180	180	180	180	180	180	180
R-squared	0.38	0.28	0.32	0.11	0.07	0.15	0.19
Wald Test Overall Significance (p-val)	0.78	0.79	0.05	0.1	0.06	0.01	0.29
Control Mean (Logbook & Scorecard)	0.67	9.21	5.17	0.08	0.21	0.88	4.25
Control SD (Logbook & Scorecard)	0.48	4.01	4.37	0.28	0.83	0.34	4.06
Control Mean (Announcement)	0.72	9.57	6.77	0.26	0.86	0.68	3.66
Control SD (Announcement)	0.45	4.12	4.84	0.44	1.7	0.47	3.91
p-values							
Scorecard vs Logbook	0.92	0.86	0.34	0.87	0.37	0.23	0.43
Scorecard only vs. Scorecard and Logbook	0.7	0.62	0.88	0.77	0.97	0.06	0.58
Scorecard only vs. Scorecard and announcement	0.96	0.15	0.72	0.56	0.19	0.28	0.37
Scorecard only vs False announcement	0.42	0.98	0.12	0.07	0.29	0.01	0.33

Controls included a set of five household-level controls: Age of household head (years), respondent is the head of the household(dummy),Household head completed primary school(dummy), number of dependents, main construction material of the walls is mud, wattle reed or other non-cement material (dummy). Robust standard errors are in parentheses and are clustered at the farmers group level. The F - test is run over all coefficients in the regression. P values of conventional omnibus test are reported of each coefficient across regressions. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Client farmers' interactions with the fee-based extension services

	Male					Female				
	Knows extension agent	Attended training	# trainings attended	Experienced problems	Reported problems	Knows extension agent	Attended training	# training attended	Experienced problems	Reported problems
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Scorecard only	-0.04 [0.08]	0.03 [0.08]	1.23*** [0.38]	0.03 [0.04]	0.01 [0.04]	0.18 [0.12]	0.25** [0.10]	0.97* [0.53]	0.02 [0.07]	-0.03 [0.06]
Logbook only	0 [0.06]	-0.07 [0.06]	0.67** [0.27]	-0.01 [0.03]	-0.02 [0.03]	0.15* [0.08]	0.14** [0.07]	0.86** [0.41]	0.03 [0.05]	0 [0.05]
False announcement	0 [0.05]	0 [0.06]	0.57* [0.31]	0 [0.03]	-0.02 [0.02]	0.02 [0.06]	0.08 [0.06]	0.34 [0.38]	0.08 [0.05]	-0.02 [0.04]
Scorecard and Logbook	0.01 [0.07]	0.01 [0.07]	0.99*** [0.30]	0.03 [0.04]	0 [0.03]	0.11 [0.11]	0.13 [0.08]	1.07** [0.46]	0.03 [0.06]	-0.02 [0.05]
Scorecard and announcement	-0.08 [0.08]	0.01 [0.08]	1.16*** [0.36]	0.01 [0.04]	-0.01 [0.04]	0.23** [0.11]	0.32*** [0.10]	0.98* [0.57]	0.04 [0.07]	-0.03 [0.07]
Observations	887	887	830	887	887	573	573	514	573	573
R-squared	0.03	0.01	0.02	0.03	0.02	0.01	0.02	0.01	0.02	0.01
Wald Test Overall Significance (P-value)	0.83	0.57	0.01	0.75	0.69	0.34	0.06	0.3	0.63	0.99
Control Mean (Logbook & Scorecard)	0.73	0.38	1.85	0.11	0.09	0.54	0.18	1.47	0.14	0.11
Control SD (Logbook & Scorecard)	0.45	0.49	2.2	0.32	0.28	0.5	0.39	2.44	0.35	0.31
Control Mean (Announcement)	0.72	0.39	2.38	0.13	0.09	0.66	0.3	2.21	0.11	0.08
Control SD (Announcement)	0.45	0.49	2.61	0.34	0.29	0.47	0.46	2.52	0.32	0.28
p-values										
Scorecard vs. Logbook	0.53	0.15	0.16	0.37	0.42	0.69	0.19	0.82	0.83	0.52
Scorecard only vs. Scorecard and Logbook	0.3	0.72	0.43	0.99	0.82	0.27	0.13	0.8	0.75	0.76
Scorecard only vs. Scorecard and announcement	0.42	0.68	0.81	0.72	0.47	0.45	0.24	0.97	0.58	0.86
Scorecard only vs False announcement	0.52	0.73	0.12	0.52	0.54	0.11	0.09	0.24	0.44	0.93

All outcomes were recorded in May-June 2013, except for "# trainings attended" which was recorded in July-August 2013, before the onset of the next season's trainings, but late enough in the season as to capture all trainings (e.g., seed selection and multiplication, and grain storage, which are delivered close to harvest). Controls included a set of five household-level controls: Age of household head (years), respondent is the head of the household(dummy),Household head completed primary school(dummy), number of dependents, main construction material of the walls is mud, wattle reed or other non-cement material (dummy). Robust standard errors are in parentheses and are clustered at the farmers group level. Wald test of Overall Significance is run over the treatment arms coefficients and the p-value is reported. The F - test as run over all coefficients in the regression. P values of conventional omnibus test are reported of each coefficient across regressions. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table 6: Client farmers' perceptions of extension provider

	Male	Female
	(1)	(2)
Scorecard only	-0.03 [0.08]	0.20* [0.10]
Logbook only	0.08 [0.06]	0.19*** [0.07]
False announcement	0 [0.06]	0.07 [0.06]
Scorecard and Logbook	0.12 [0.07]	0.19** [0.09]
Scorecard and announcement	-0.06 [0.09]	0.23** [0.10]
Observations	875	561
R-squared	0.04	0.06
Wald Test Overall Significance (P-value)	0.13	0.14
Control Mean (Logbook & Scorecard)	0.43	0.35
Control SD (Logbook & Scorecard)	0.48	0.46
Control Mean (Announcement)	0.51	0.49
Control SD (Announcement)	0.48	0.48
<i>p-values</i>		
Scorecard vs. Logbook	0.12	0.93
Scorecard only vs. Scorecard and Logbook	0.01	0.95
Scorecard only vs. Scorecard and announcement	0.51	0.57
Scorecard only vs False announcement	0.67	0.19

Controls included a set of five household-level controls: Age of household head (years), respondent is the head of the household(dummy),Household head completed primary school(dummy), number of dependents, main construction material of the walls is mud, wattle reed or other non-cement material (dummy). Robust standard errors are in parentheses and are clustered at the farmer group level. Wald test of Overall Significance is run over the treatment arms coefficients and the p-value is reported. The F - test as run over all coefficients in the regression. P values of conventional omnibus test are reported of each coefficient across regressions. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Appendix

Table A1: Farmers' knowledge score

	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
	Fertilizer application	Main Crop	All topics	Fertilizer application	Main Crop	All topics
Scorecard only	-13.27*	-4.26	-0.98	-10.06	-4.42	2.18
	[8.00]	[5.65]	[3.12]	[10.99]	[6.46]	[4.51]
Logbook only	-4.69	-0.09	0.13	9.42	-3.78	-1.72
	[6.23]	[4.28]	[2.44]	[8.71]	[4.40]	[2.57]
False announcement	2.28	-6.70*	0.14	2.76	-6.69	-3.62
	[6.05]	[3.60]	[1.85]	[6.97]	[4.86]	[2.94]
Scorecard and Logbook	-1.46	-2.7	0.26	1.76	-0.85	-0.12
	[7.13]	[4.92]	[2.78]	[10.21]	[5.43]	[3.25]
Scorecard and announcement	-12.13	-0.94	2.88	5.07	-7.71	1.9
	[7.78]	[5.08]	[2.91]	[12.26]	[6.99]	[4.44]
Observations	782	594	830	506	379	514
R-squared	0.03	0.03	0.02	0.08	0.11	0.08
Wald Test Overall Significance (P-value)	0.22	0.46	0.33	0.06	0.52	0.45
Control Mean (Logbook & Scorecard)	48.15	30.77	40.22	32.69	31.6	38.22
Control Mean (Announcement)	48.96	28.47	20.98	46.32	27.36	15.62
Control SD (Logbook & Scorecard)	47.62	32.8	41.73	38.11	33.86	41.97
Control SD (Announcement)	47.85	29.86	19.87	46.99	28.24	20.06
p-values						
Scorecard vs Logbook	0.22	0.41	0.64	0.02	0.92	0.37
Scorecard only v.s Scorecard and Logbook	0.02	0.69	0.5	0.15	0.52	0.52
Scorecard only v.s Scorecard and announcement	0.84	0.37	0.06	0.03	0.47	0.92
Scorecard only vs False announcement	0.06	0.67	0.71	0.23	0.75	0.18

Controls included a set of five household-level controls: Age of household head (years), respondent is the head of the household(dummy),Household head completed primary school(dummy), number of dependents, main construction material of the walls is mud, wattle reed or other non-cement material (dummy). Robust standard errors are in parentheses and are clustered at the farmer group level. Main crop includes knowledge score for maize and beans. Overall is the average of the individual knowledge scores. Fertilizer application: Correct amount of fertilizer (DAP or NPK) needed for crops (std by number of crops planted). Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A2: Quantity of extension services delivered at the group level

	% loan repaid in 2013	Attended training (max)	# trainings attended (average)
	(1)	(2)	(3)
Scorecard only	0.08 [2.98]	0 [0.09]	0.39 [0.26]
Logbook only	-0.08 [1.93]	-0.06 [0.07]	0.17 [0.18]
False announcement	2.2 [1.75]	0.03 [0.07]	0.14 [0.21]
Scorecard and Logbook	1.06 [2.46]	0.06 [0.08]	0.3 [0.20]
Scorecard and announcement	0.07 [2.91]	-0.04 [0.09]	0.49* [0.28]
Observations	180	180	180
R-squared	0.52	0.05	0.12
Wald Test Overall Significance (P-value)	0.83	0.47	0.53
Control Mean (Logbook & Scorecard)	87.5	0.92	0.79
Control SD (Logbook & Scorecard)	13.13	0.28	0.55
Control Mean (Announcement)	89.08	0.91	1.05
Control SD (Announcement)	10.96	0.29	0.92
<i>p-values</i>			
Scorecard vs Logbook	0.59	0.22	0.67
Scorecard only vs. Scorecard and Logbook	0.59	0.22	0.67
Scorecard only vs. Scorecard and announcement	0.99	0.32	0.6
Scorecard only vs False announcement	0.43	0.72	0.36

Controls included a set of five household-level controls: Age of household head (years), respondent is the head of the household(dummy),Household head completed primary school(dummy), number of dependents, main construction material of the walls is mud, wattle reed or other non-cement material (dummy). Robust standard errors are in parentheses and are clustered at the farmer group level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A3: Client farmers' attendance (by training)

Panel A: Male Farmers

	Pest management	Compost making	Planting	Fertilizer application	Nurseries	Kitchen garden	Seed selection	Grain storage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Scorecard only	0.11*	0.09	0.21**	0.08	0.21***	0.15***	0.23***	0.14***
	[0.06]	[0.08]	[0.08]	[0.08]	[0.05]	[0.06]	[0.06]	[0.05]
Logbook only	0.11**	0.01	0.13**	0.12*	0.10***	0.07	0.11***	0.03
	[0.05]	[0.06]	[0.06]	[0.06]	[0.04]	[0.04]	[0.04]	[0.04]
False announcement	0.04	0.10*	0.08	0.01	0.09*	0.06	0.12***	0.07*
	[0.05]	[0.06]	[0.07]	[0.06]	[0.05]	[0.05]	[0.04]	[0.04]
Scorecard and Logbook	0.13**	0.11*	0.14**	0.11	0.14***	0.09**	0.16***	0.11***
	[0.05]	[0.07]	[0.07]	[0.07]	[0.04]	[0.05]	[0.05]	[0.04]
Scorecard and announcement	0.16**	0.1	0.22***	0.1	0.18***	0.16***	0.18***	0.06
	[0.06]	[0.07]	[0.07]	[0.08]	[0.05]	[0.05]	[0.06]	[0.05]
Observations	830	830	830	830	830	830	830	830
R-squared	0.02	0.02	0.01	0	0.04	0.01	0.05	0.02
Wald Test Overall Significance (P-value)	0.06	0.38	0.06	0.5	0.01	0.1	0	0.05
Control Mean (Logbook & Scorecard)	0.21	0.42	0.34	0.36	0.09	0.18	0.13	0.11
Control Mean (Announcement)	0.41	0.5	0.48	0.48	0.28	0.39	0.34	0.32
Control SD (Logbook & Scorecard)	0.29	0.43	0.43	0.46	0.16	0.21	0.23	0.17
Control SD (Announcement)	0.46	0.5	0.5	0.5	0.36	0.41	0.42	0.38
<i>p-values</i>								
Scorecard vs. Logbook	0.99	0.19	0.32	0.66	0.04	0.16	0.03	0.04
Scorecard only vs. Scorecard and Logbook	0.76	0.75	0.28	0.66	0.1	0.2	0.11	0.45
Scorecard only vs. Scorecard and announcement	0.25	0.81	0.77	0.82	0.38	0.9	0.22	0.06
Scorecard only vs. False announcement	0.26	0.88	0.15	0.35	0.03	0.16	0.06	0.19

Controls included a set of five household-level controls: Age of household head (years), respondent is the head of the household(dummy),Household head completed primary school(dummy), number of dependents, main construction material of the walls is mud, wattle reed or other non-cement material (dummy). Robust standard errors are in parentheses and are clustered at the farmers group level. Wald test of Overall Significance is run over the treatment arms coefficients and the p-value is reported. The F - test as run over all coefficients in the regression. P values of conventional omnibus test are reported of each coefficient across regressions. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A3: Client farmers' attendance (by training)

Panel B: Female Farmers

	Pest management	Compost making	Planting	Fertilizer application	Nurseries	Kitchen garden	Seed selection	Grain storage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Scorecard only	0.11 [0.09]	0.27*** [0.10]	0.33*** [0.10]	0.15 [0.11]	0 [0.07]	0.06 [0.08]	0.03 [0.08]	0.01 [0.07]
Logbook only	0.15** [0.07]	0.17** [0.07]	0.14* [0.08]	0.19** [0.08]	0.08 [0.05]	0.10* [0.06]	0.01 [0.06]	0.03 [0.06]
False announcement	0.07 [0.06]	0.06 [0.06]	0.06 [0.07]	-0.01 [0.06]	0.05 [0.06]	0.03 [0.06]	0.03 [0.06]	0.04 [0.05]
Scorecard and Logbook	0.21** [0.09]	0.18** [0.08]	0.24** [0.09]	0.2** [0.09]	0.1 [0.07]	0.07 [0.07]	0.05 [0.07]	0.02 [0.06]
Scorecard and announcement	0.11 [0.10]	0.31*** [0.10]	0.33*** [0.10]	0.14 [0.12]	-0.01 [0.08]	0.04 [0.09]	0.02 [0.10]	0.03 [0.08]
Observations	514	514	514	514	514	514	514	514
R-squared	0	0.03	0.03	0.02	0.01	-0.02	0	-0.02
Wald Test Overall Significance (P-value)	0.17	0.05	0.01	0.24	0.22	0.5	0.98	0.91
Control Mean (Logbook & Scorecard)	0.21	0.26	0.25	0.25	0.09	0.14	0.18	0.11
Control Mean (Announcement)	0.41	0.44	0.43	0.43	0.29	0.35	0.38	0.31
Control SD (Logbook & Scorecard)	0.3	0.43	0.41	0.41	0.14	0.2	0.21	0.11
Control SD (Announcement)	0.46	0.5	0.49	0.49	0.35	0.4	0.41	0.31
p-values								
Scorecard vs Logbook	0.67	0.28	0.02	0.64	0.21	0.57	0.68	0.76
Scorecard only vs. Scorecard and Logbook	0.13	0.26	0.13	0.53	0.05	0.9	0.86	0.84
Scorecard only vs. Scorecard and announcement	0.98	0.48	0.98	0.93	0.97	0.68	0.8	0.71
Scorecard only vs. False announcement	0.6	0.04	0	0.16	0.44	0.65	0.96	0.65

Controls included a set of five household-level controls: Age of household head (years), respondent is the head of the household(dummy), Household head completed primary school(dummy), number of dependents, main construction material of the walls is mud, wattle reed or other non-cement material (dummy). Robust standard errors are in parentheses and are clustered at the farmers group level. Wald test of Overall Significance is run over the treatment arms coefficients and the p-value is reported. The F - test as run over all coefficients in the regression. P values of conventional omnibus test are reported of each coefficient across regressions. Significance levels: *** p<0.01, ** p<0.05, * p<0.1