"List Randomization for Sensitive Behavior:

An Application for Measuring Use of Loan Proceeds

Dean Karlan
Yale University
Innovations for Poverty Action
M.I.T. Jameel Poverty Action Lab

Jonathan Zinman
Dartmouth College
Innovations for Poverty Action
M.I.T. Jameel Poverty Action Lab

August 2011

Policymakers and microfinance institutions (MFIs) often claim to target poor entrepreneurs who then invest loan proceeds in their businesses. Typically in non-research settings these claims are assessed using readily available but unverified self-reports from client loan applications. Alternatively, independent surveyors could directly elicit how borrowers spent their loan proceeds. That too, however, could suffer from deliberate misreporting. We use data from the Peru and the Philippines in which independent surveyors elicited loan use both directly (i.e., by asking how individuals spent their loan proceeds) and indirectly (i.e., through a list-randomization technique that allows individuals to hide their answer from the surveyor). We find that direct elicitation under-reports the non-enterprise uses of loan proceeds.

> "Microfinance is a proven tool for fighting poverty on a large scale. It provides very small loans, or micro-loans, to poor people, mostly women, to start or expand very small, self-sufficient businesses."
>
> *- Quote from website of Grameen Foundation USA, a leading microcredit organization*

Introduction

Policymakers place increasing emphasis on expanding outreach to poorer (potential) entrepreneurs, and microfinance institutions (MFIs) often claim to target poor entrepreneurs who then invest loan proceeds in growing their businesses. Typically these claims are evaluated using readily available but unverified self-reports from client loan applications. We examine whether MFIs and third-parties can rely on client self-reports to learn how individuals spent the loan proceeds.

If there is any incentive to lie, self-reports from clients to MFIs are likely to be biased in whichever direction serves the interests of the clients. Even if clients are assured that their answers will not affect their loan eligibility, respondents may lie if they do not trust the surveyors' (or loan officers') guarantees or if they wish to project a socially desirable image.

Note that there is a more important impact question that we do not address: what expenditures or investments were made that would not have been made had that lender not made the loan. To answer this question, one needs a measure of the counterfactual: what would have happened without a loan. In this paper we focus strictly on borrower reports of how they spent loan proceeds, and how they report this differently whether asked by the lender, a surveyor directly, or a surveyor indirectly in a way that allows the respondent to conceal their answer.

We report here on two mini-studies on borrower "use of funds" that are part of larger ongoing studies with MFIs. The first, with Arariwa in Peru, uses a survey technique called "list randomization" (explained below), to assess whether individuals feel compelled to underreport using loan proceeds for consumption, rather than investment. The presumption is that if individuals underreport using funds for consumption *to an independent surveyor*, then they will likely also underreport the same if asked by a lender who emphasizes using loans for entrepreneurial purposes. The second study, with First Macro Bank in the Philippines, examines the key underreporting hypothesis more directly, by comparing reports on non-business uses across three elicitation methods: direct questioning by the bank, direct questioning by the surveyor, and list randomization presented by the surveyor.

Design, Data, and Results

List Randomization

Researchers in the social sciences have developed a variety of techniques that attempt to elicit truthful responses to sensitive questions. One approach includes direct methods such as matching

the gender of surveyors and respondents, using forgiving language, using unfolding financial brackets, and collecting data in private. A second approach involves using indirect methods such as the randomized response technique, the bogus pipeline, and the list randomization technique used here.[1]

List randomization, also known as the item count or unmatched count technique, provides a simple way for respondents to report on sensitive behavior without allowing the researcher or surveyors to identify individual responses. To employ this technique, half of the survey respondents are randomly selected to receive a short list of statements (in our case a list of business investments) and asked to report how many, but not which, statements are true. The other half of the survey respondents are presented with the same list of statements and one key additional statement designed to capture sensitive behavior (in our case non-business investment or a type of consumption). By subtracting the mean number of true statements in the first group from the mean number of true statements in the second group, researchers can estimate the proportion of the sample that engages in the sensitive behavior.

Several studies suggest that the randomized list technique can yield more accurate responses to sensitive survey questions compared to the direct reporting method. Across 48 comparisons of direct report and list randomization, one meta-analysis found that 63% of the estimates for socially undesirable behavior were significantly larger when elicited through list randomization (Holbrook and Krosnick 2009). A more limited meta-analysis found that while the list randomization estimates of socially undesirable behavior were generally larger, particularly for studies using undergraduate samples, the overall difference was not significant (Tourangeau 2007).

To validate the method as a means to elicit information about specifically sensitive behavior, some studies have more precisely estimated the effectiveness of the technique by comparing direct report to list randomization for both sensitive and non-sensitive questions. Tsuchiya et al. (2007) finds that the technique results in a significantly higher proportion of a sample admitting to shoplifting, whereas the difference between methods in estimates of blood donation is insignificant. Similarly, LaBrie and Earleywine (2000) finds that list randomization results in a higher proportion of undergraduate students admitting to having unprotected sex, whereas there was no significant difference for drinking alcohol, which presumably has less stigma.

[1] The randomized response technique was first developed in 1965 by Stanley Warner as a process in which a randomizing device such as a spinner would select one of two statements about a sensitive topic. The spinner would select one statement with known probability p and the other statement with probability 1-p. The respondent would then inform the surveyor whether or not she agreed with the selected statement, without disclosing which statement was selected by the spinner. Other indirect methods include the unrelated question technique, the forced alternative technique, and the bogus pipeline technique. In the unrelated question technique, respondents are asked to answer "yes" or "no" to one of two randomly selected questions: the sensitive question or a question with a known probability of a "yes" answer. In the forced alternative technique, the respondent is presented with a sensitive question and then uses a randomizing device to determine whether to respond "yes", "no", or to present her true response. The bogus pipeline technique tells respondents they are being monitored by a lie detector.

One challenge of the list randomization method lies in the selection of the non-key items in the list. In order to reduce variance, the values of non-key list items should have as little variance as possible. That is, the non-key items should describe relatively innocuous behaviors that almost everyone has done, or almost everyone has not done. But if the items represent behaviors that pose no variation across people, the respondent may not feel confident that his or her answer about the behavior in question would be anonymous. As a result of this dilemma, list randomization often produces results that are too high in variance to be statistically significant, especially if the behavior of interest is low prevalence (which it often is, since high prevalence behaviors are typically not that sensitive in the first place) (Droitcour et al. 1991). We suspect loan use for consumption purposes to be common enough to warrant the application of this method here.

Another critique of list randomization lies in how the method is presented to respondents. Giving a more detailed explanation of the technique reassures respondents that their answers will remain anonymous, and therefore results in higher reports of the sensitive behavior (Ahart and Sackett 2004). There is no evidence, however, that the number of non-key items in the list affects the difference between the direct response and list randomization estimates, implying that we can gain relatively accurate estimates from lists of three, four, five, or six items (Tsuchiya et al. 2007). In comparison to other indirect methods, list randomization is often more simple to administer (both for surveyors and respondents) and effective (Droitcour et al. 1991).

Another drawback of the list randomization technique is that it generates only aggregate information. While it reveals information about the rate of presence of the sensitive behavior in a population, the anonymity of the method makes it impossible to examine the relationship between the behavior and individual characteristics due to the anonymity of the method. Breaking down base rate analysis by subgroups defined by another individual measure can allow for more subtle exploration of the relationship between the sensitive behavior and individual characteristics (Ahart and Sackett 2004).

Study 1: Arariwa MFI, Peru

Our first mini-study compares borrower reports of loan uses from two different elicitation methods implemented by *surveyors*: direct questioning versus list randomization. Prior to evaluating the use of video and radio as a means for financial education, 1650 MFI clients were surveyed in Cuzco, Peru. The lending institution, Arariwa, provides microcredit for business purposes to approximately 20,000 low-income households in southeastern Peru. Arariwa emphasizes that loans should be used for business, and requires the borrower to state what the loan will be used for when they apply. However, there is no policy of explicitly monitoring the use of the cash proceeds from loan disbursal. As part of the baseline survey, Arariwa clients were asked questions related to their personal finances and education. Surveyors were not affiliated with any MFI and informed survey respondents that their responses would not be shared with

anyone other than researchers studying how entrepreneurs that are Arariwa clients manage their household finances.

All respondents were asked to report their loan uses through direct report and list randomization techniques. For the direct report, respondents were asked to list up to five loans that they had taken out in the past 12 months, by loan source and amount. They were then asked, "Which need or which needs did you cover with this loan?" and allowed to list up to three uses for each loan.[2] Though respondents were not prompted with categories, surveyors matched uses against one of 18 possibilities (Table 1). Note that this thus is not a perfect match to the list randomization question; the second study, detailed below, addresses this flaw by matching the questions more precisely. After eight more questions related to personal finances, respondents were presented with the list randomization module.

In surveyor training, we explained that the list randomization intended to ask private questions in an anonymous fashion. The surveyors did not understand the details of the calculation, but they understood that the process generated anonymity and the importance of making this anonymity clear to clients. They also understood the importance of maintaining the random assignment to treatment groups.

Prior to beginning the list randomization, surveyors were instructed to demonstrate the technique using an example. Surveyors were provided with five innocuous statements printed on a piece of paper with a clear clipboard placed over the sheet. Respondents were handed the clipboards and asked to use a white board marker to put check marks next to statements that are true for them. Next, respondents were instructed to count the number of true statements before erasing their check marks, returning the clipboard, and reporting the total count. After confirming that the clients understood the anonymity ensured by the process, surveyors began the list randomization module.

Clients were randomly selected to be presented with one of four possible groups of three to six statements.[3] All clients received the following three statements: "I used part of my Arariwa loan to buy *merchandise* for my economic activity", "I used part of my Arariwa loan to buy *equipment* for my economic activity" and "I *shared* my loan with another person". Clients in group A (n=408) only received these statements. Clients in group B (n=414) additionally received the following statement: "I used at least a quarter of my Arariwa loan on *household*

[2] Only 1.5% of the sample listed five loans, implying that respondents were not limited by the survey options to underreport loans. Similarly, 2.4% of all loans and 2.1% of loans identified as "Loan from Arariwa" or "Loan from Communal Bank (facilitated b Arariwa)" had three uses, implying that the three-use maximum was not binding for most respondents.

[3] **The randomization was stratified by lending group. A subset of clients were randomly selected to be surveyed, and if an individual was not found then there was a replacement list, randomly ordered, of individuals to survey. Any replacement individual was assigned to the same list randomization treatment as the original target respondent.**

items, such as food, a TV, a radio, etc." Group C (n=388) received the four previous statements, and the statement, "I used at least a quarter of my Arariwa loan to pay for my family's *medical expenses*." Group D (n=401) received the previous five statements and the statement, "I used at least a quarter of my Arariwa loan to pay for my family's *educational expenses*." By subtracting the mean number of true statements for group A from the mean number of true statements for group B, we get the proportion of clients that used a quarter of their loan for household items. We similarly subtract B from C and C from D to get the proportions of clients using their loans on education or medical expenses.

In order to compare estimates, we match the loan uses from direct report to those from list randomization. Since the direct report question allows clients to list up to five loans from any source, we limit the sample to only include Arariwa loans or communal loans facilitated by Arariwa.[4] Due to cultural norms and surveyor training, "household items" is best approximated by the direct report responses that are classified as "consumption good", "purchase clothing or shoes", and "other consumption need."

[4] In piloting the survey, clients did not seem to differentiate between loans directly from Arariwa and loans from the savings accumulated by peers in village banks organized by Arariwa.

Table 1:
Loan Uses from Direct Response Question
from ICT Financial Literacy Project in Peru

Use	Mean	Standard Error
Use, by Category		
Any Production [Responses (1) - (8) or (-666)]	0.758	0.011
Any Consumption [Responses (9) - (17) or (-667)]	0.300	0.011
Household Item [Responses (13), (15), or (-667)]	0.077	0.007
Use, by Specific Response		
(1) Purchase land	0.022	0.004
(2) Purchase equipment	0.068	0.006
(3) Agricultural inputs (fertilizer, pesticide, etc)	0.051	0.005
(4) Purchase animals	0.179	0.009
(5) Animal husbandry inputs (fodder, medicines, etc)	0.021	0.004
(6) Raw materials	0.090	0.007
(7) Purchase merchandise	0.411	0.012
(8) Purchase of assets to enable a shop or office	0.021	0.004
(9) Education	0.072	0.006
(10) Health	0.022	0.004
(11) Ceremonies(weddings, funerals, etc)	0.004	0.002
(12) Purchase of vehicles	0.020	0.003
(13) Consumption goods	0.052	0.005
(14) To pay off another loan	0.042	0.005
(15) Purchase clothing and shoes	0.008	0.002
(16) Travel	0.008	0.002
(17) Home improvement	0.067	0.006
(-666) Other productive need	0.061	0.006
(-667) Other consumption need	0.020	0.003

N = 1650. An individual use = 1 if it is listed as any of three possible uses across any of five possible loans. Only loans identified as "Loan from Arariwa" or "Loan from Communal Bank (facilitated by Arariwa)" are included. Only 2.1% of those loans had three uses, implying that the three-use maximum was not binding for most respondents.

Table 2 demonstrates a striking contrast in results between direct questioning and list randomization. Direct questioning reveals only 7.7% of the sample volunteering household items as a use for any of their Arariwa or Arariwa-facilitated loans. In comparison, the list randomization technique suggests that 31.3% of the sample used at least a quarter of their Arariwa loans on household items. Similarly, 2.2% of the sample volunteered a health related loan use through direct questioning, whereas list randomization resulted in an estimate of 23.1% of the sample using at least a quarter of their loan amounts on medical expenses. Finally, the proportion for clients using loans for educational expenses is 7.1% through direct questioning, but 33.2% through list randomization. Z-tests of proportions indicate that each of these three differences is statistically significant.

Although the magnitude of the estimated underreporting here is large, it is consistent with results from other studies. For example, Karlan and Zinman (2007) look at the "cash loan" market in South Africa, and compare self-reports on loans with administrative data. They find that nearly 50% of respondents lie about their borrowing activity.

There are several reasons why list randomization might produce such different, and higher, estimates of loan uses than direct report. Asking clients to do direct report first and list randomization second biases the results. Future research could test this by randomizing the order of direct report and lists. Another issue is whether list randomization reduces lying, and/or facilitates recall. Future research could test this by comparing direct reports versus list randomization on topics not likely to be sensitive (e.g., asking about using microloan proceeds for business expenses), and/or by testing how prompting specific categories changes responses in direct elicitation.

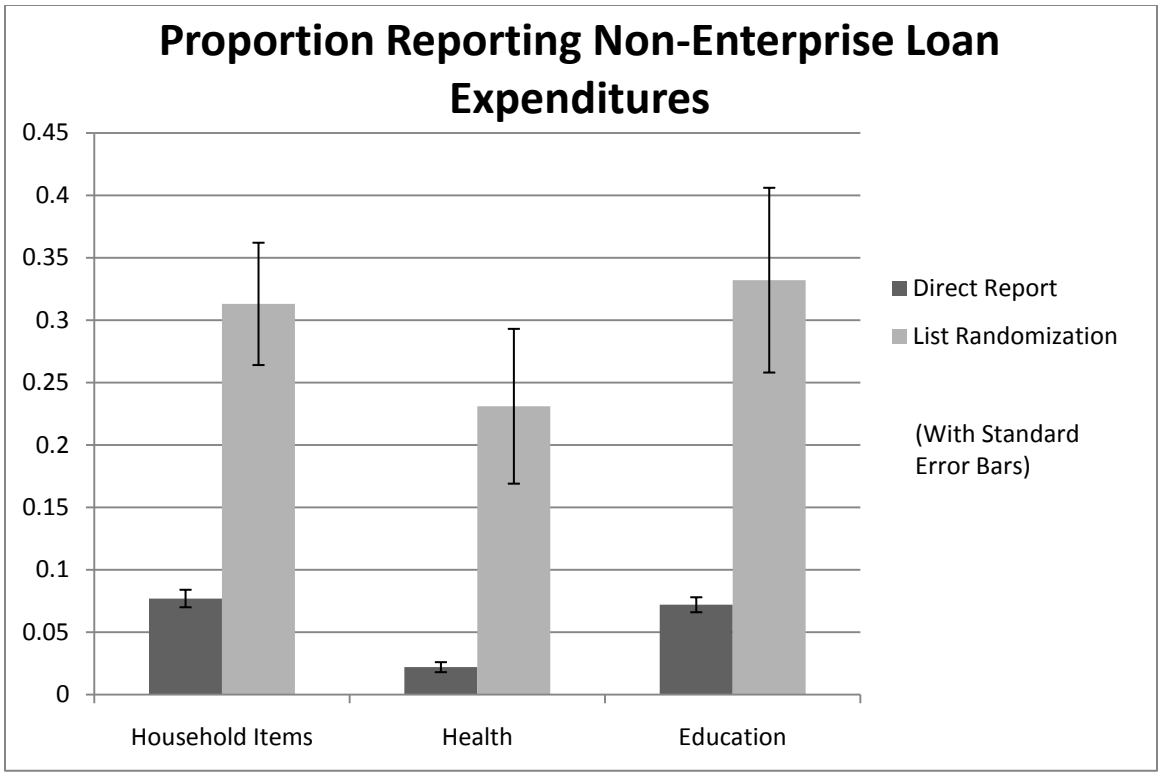

Proportion Reporting Non-Enterprise Loan Expenditures
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Household Items
Health
Education
Direct Report
List Randomization
(With Standard Error Bars)

Table 2:
Comparison of Direct Report and List Randomization Estimates
from ICT Financial Literacy Project in Peru

Loan Use:	Household Items (1)	Health (2)	Education (3)
Direct Report			
Proportion reporting this use	0.077	0.022	0.072
SE	(0.007)	(0.004)	(0.006)
N	1650	1650	1650
List Randomization			
Mean of "Yes" Responses for Short List	1.213	1.527	1.758
SE	(0.031)	(0.038)	(0.049)
N	408	414	388
Mean of "Yes" Responses for Long List	1.527	1.758	2.090
SE	(0.038)	(0.049)	(0.055)
N	414	388	401
Difference (Proportion reporting this use)	0.313	0.231	0.332
SE of Difference	(0.049)	(0.062)	(0.074)
p-value from ttest	0.000	0.000	0.000
N	822	802	789
Comparison of Direct Report & List Randomization			
List Randomization minus Direct Report	0.236***	0.209***	0.261***
Z-test statistic for difference in proportions	4.752	3.386	3.512

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in parentheses. Direct report question allows up to three uses to be reported for each of five loans. Only loans directly from Arariwa or facilitated by Arariwa are included. "Household items" question from list randomization is matched to the following direct report options: "consumption goods", "purchase clothing or shoes", and "other consumption need". List randomization questions required that over 1/4 of the loan was used for the specified purpose, whereas the direct report question did not have a lower bound on proportion of loan used.

Study 2: FMB, The Philippines

In our second mini-study, clients at three banks in the Philippines were subjected to two questions at four different times during their loan cycle. The questions aimed to get at the truth behind two statements: (1) "I used 2500 pesos or more of my loan to pay down other debt" and (2) "I used 5,000 pesos or more of my loan on any single transaction for my household". (Respondents were asked to consider the statements with regard to their most recent loan.) The four steps of elicitation are detailed in Table 3. First, credit officers presented the questions to clients on their loan application. Credit officers then presented the questions again when clients went to make their first loan repayment. These two instances allow us to see how answers change before and after the loan was granted.

Surveyors then visited clients, on average, two weeks after the client was granted a loan from one of the participating banks.[5] The surveyor asked them to participate in a survey about "Health and Financial Services." Respondents had no reason to believe that the surveyors had any connection to the bank. The first few questions asked about health attitudes and behaviors so that clients would not think that the surveys were coming directly from the bank. Surveyors had no information about the three participating banks. Surveyors then asked the clients the two questions explicitly. The difference between the responses during the first repayment and the explicit questions from the surveyor allows us to see how responses change when clients think the bank may be monitoring their answers. At that time, surveyors also presented the questions indirectly using list randomization.

Clients were asked two sets of list randomization questions. These questions allowed us to estimate the proportion of true answers to the two statements. Each client randomly received one of four surveys. All surveys contained the following four statements in the first question: "I have visited a hospital or clinic in the last six months," "I have more than three siblings," "I have purchased some type of insurance in the past five years," and "My household owns an air conditioner." The second and fourth surveys had "I used 2,500 pesos or more of my loan to pay down other debt" as the fifth question.

Similarly the second set of list randomization questions included the following four statements on all surveys: "I have a washing machine in my home," "I am originally from this city," "I have completed one year or more of formal education post-high school," and "My household owns a computer." The third and fourth surveys also include the statement "I used 5,000 pesos or more of my loan on any single transaction for my household." In this case, the questions used in list randomization were exactly the same as those used in direct elicitation, so any differences in results can be attributed purely to the method and not the content of the question.

[5] We also surveyed some clients who were denied loans from the bank. These clients were part of a larger overall experiment in which we randomized loan decisions on marginally creditworthy clients. Future work will look at the difference in expenditures amongst randomized clients, thus taking into account the fungability of money.

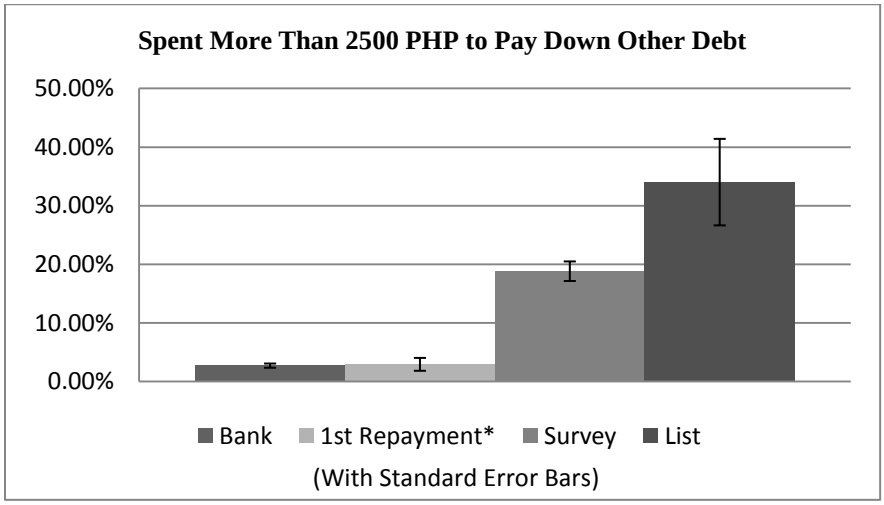
Survey one was administered to 58 people in our sample, survey two was administered to 77 people, survey three was administered to 59 people and the final survey was administered to 66 people. Comparing results from the explicit question by the surveyor to results from the list randomization will demonstrate how responses change when clients believe their answers are truly private, even from the surveyor[6]

Table 3

STEP	WHO?	WHEN?	HOW?
A	Credit officer	Bank application	Direct question
B	Credit officer	1st repayment	Direct question
C	Surveyor	2 weeks after loan disbursal	Direct question (no prompt)
D	Surveyor	2 weeks after loan disbursal	List randomization

Chart 1 shows the results for Question 1, and Chart 2 shows the results for Question 2. When the credit officers asked clients directly, for both questions, fewer than 4% of clients admitted that the statements were true (note that due to compliance issues, our sample size is smaller for Step B, hence the larger standard error). When surveyors asked the questions, response rates jumped up, and when surveyors asked the questions using list randomization, response rates rose even further.

Chart 1



[6] We had one final measure of whether or not clients actually used their loans to pay down debt or for household expenditures. In the study there was a treatment group, which received loans from FMB, and a control group, which did not. In both groups, surveys were used to measure expenditure. One survey was conducted two weeks after the first repayment, at the same time that the direct questions were presented. The second survey was conducted much later. Treatment groups showed higher levels of debt repayment and household expenditure than control groups.

Chart 2

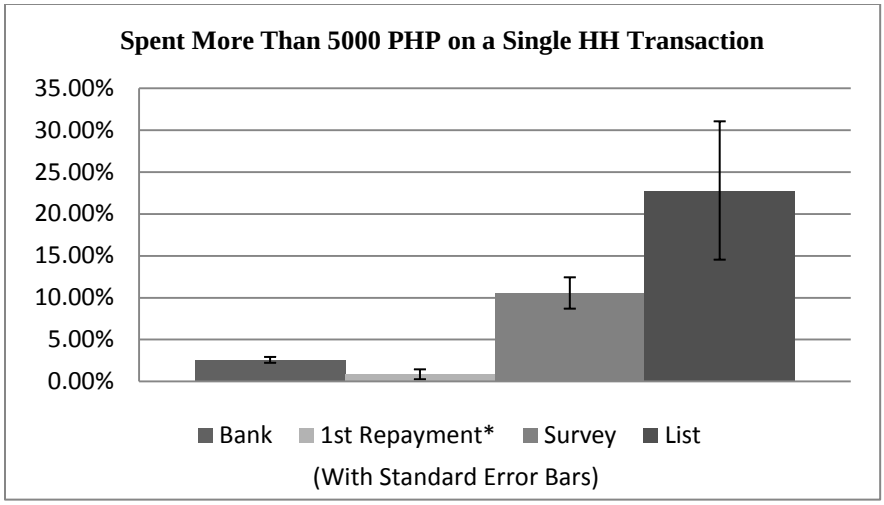

Spent More Than 5000 PHP on a Single HH Transaction
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
Bank
1st Repayment*
Survey
List
(With Standard Error Bars)

Table 4:
Comparison of Various Responses
from First Macro Bank in the Philippines

Loan Use:		Spent More Than 2500 PHP to Pay Down Other Debt (1)	Spent More Than 5000 PHP on a Single HH Transaction (2)
Bank Responses			
Proportion reporting this use		0.0272	0.0256
	SE	(0.0036)	(0.0035)
	N	2061	2067
First Repayment Responses			
Proportion reporting this use		0.0294	0.0084
	SE	(0.0110)	(0.0059)
	N	238	238
Survey Responses			
Proportion reporting this use		0.1883	0.1055
	SE	(0.0167)	(0.0187)
	N	749	749
List Randomization			
Mean of "Yes" Responses for Short List		1.574	2.194
	SE	(0.052)	(0.053)
	N	364	392
Mean of "Yes" Responses for Long List		1.915	2.422
	SE	(0.052)	(0.064)
	N	386	358
Difference (Proportion reporting this use)		0.340	0.228
	SE of Difference	(0.074)	(0.083)
	p-value from ttest	0.000	0.003
	N	750	750

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in parentheses. Direct report question allows up to three uses to be reported for each of five loans. Only loans directly from Arariwa or facilitated by Arariwa are included. "Household items" question from list randomization is matched to the following direct report options: "consumption goods", "purchase clothing or shoes", and "other consumption need". List randomization questions required that over 1/4 of the loan was used for the specified purpose, whereas the direct report question did not have a lower bound on proportion of loan used.

Conclusion

Data on the loan uses of (potential) microfinance clients are important inputs into business strategy and policy evaluation. We have highlighted some challenges in eliciting accurate measures, presented some evidence suggesting that data collected by different methods produces different inferences, and highlighted several directions for further research.

On a substantive, policy level, we learn a key lesson, and suggest a second for further research. First, we show that clients demonstrate major biases in self-reports on the use of microcredit loan proceeds. The MFI community often claims and advertises a strict focus on enterprise investment. Here we find evidence of substantial perception of consumption uses by clients. More to the point, we find that microcredit clients significantly overreport enterprise investments and underreport consumption uses to credit officers, and *even to independent surveyors*. Second, we also see using the list randomization technique as an interesting tool to determine under what conditions people deliberately misreport information. This is useful not just methodologically, but also in that it reveals information about social norms that could be interesting in its own right. Such analysis clearly could be heterogeneous, and thus using this tool on larger sample sizes and other topics could provide insightful.

Note that even accurate honest self-reports on loan uses have their limitations, and should not be considered a measure of how the proceeds were really used. Money is fungible, and hence observing the mechanical deployment of loan proceeds does not identify answers to what are typically the greater questions of interest: how does credit access change actual expenditures shortly after loan disbursal, whether on investment or consumption goods? Identifying such impacts requires data on a valid comparison group of would-be borrowers that did not get a loan for some exogenous reason. Future research from this second study, which generates such exogenous variation by introducing some randomness into bank decisions on marginal applications, will shed insight into this question.

References:

Ahart, Allison M. and Paul R.Sackett. 2004. "A New Method of Examining Relationships between Individual Difference Measures and Sensitive Behavior Criteria: Evaluating the Unmatched Count Technique." Organizational Research Methods 7:101-114.

Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, and Trena M. Ezzati. "The Item Count Technique as a Method of Indirect Questioning: a Review of its Development and a Case Study Application," in: Beimer, P.B., Groves, R.M., Lyberg L.E., Mathiowetz N.A., Sudman S. (Eds.), *Measurement Errors in Surveys*. John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 185-211.

Holbrook, Allyson L., and Jon A. Krosnick. 2009. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique." Public Opinion Quarterly (in press).

Karlan, Dean and Jonathan Zinman. 2007. "Lying about Borrowing." Journal of the European Economic Association (Papers and Proceedings) 6: 510-521.

LaBrie, Joseph W., and Mitchell Earleywine. 2000. "Sexual Risk Behaviors and Alcohol: Higher Base Rates Revealed Using the Unmatched-Count Technique. Journal of Sex Research 37:321–26.

Tsuchiya, Takahiro, Yoko Hirai, and Shigeru Ono. 2007. "A Study of the Properties of the Item Count Technique." Public Opinion Quarterly 71:253-272.

Tourangeau, Roger and Ting Yan. 2007. "Sensitive Questions in Surveys." Psychological Bulletin 133:859-883.