Survey Research Methods (2022)
Vol. 16, No. 2, pp. 165-206
doi:10.18148/srm/2022.v16i2.7651
European Survey Research Association

Boosting Survey Response Rates by Announcing Undefined Lottery Prizes in Invitation Email Subject Lines: Evidence from a Global Randomized Controlled Trial

Syedah Ahmad
Faculty of Economics and Business
University of Groningen
The Netherlands

Robert Lensink
Faculty of Economics and Business
University of Groningen
The Netherlands

Annika Mueller
Faculty of Economics and Business
University of Groningen
The Netherlands

To test whether announcing undefined lottery prizes (AULP) in the subject line of survey invitation emails influences contact and response rates, we conduct a unique, randomized, controlled trial using a multicultural, multinational sample of 5,128 key staff members of microfinance institutions from 124 countries, half of whom randomly receive the AULP treatment. By applying the leverage-salience theory of survey participation, proposed by Groves, Singer, and Corning (2000), we formulate three main hypotheses and establish three main findings. (1) In line with AULP increasing the salience of the underlying lottery incentive, we find that on average, contact and response rates are significantly higher in the AULP treatment group. (2) In line with respondents in Organization of Islamic Cooperation (OIC) countries assigning a more negative leverage to the underlying lottery, reflecting more negative socio-cultural attitudes toward gambling stemming from Islamic religious norms, we find the treatment effect of the salience-inducing AULP treatment to be lower among respondents in OIC countries than in non-OIC countries. (3) Consistent with translation provision into a local language enhancing the salience of the underlying lottery further, we find that the positive effect of AULP is accentuated for the subgroup of non-OIC countries that received translations, and the effect of AULP is even lower for the subgroup of OIC countries that received translations, than for those which did not. The results offer no indication that AULP leads to non-response bias or other adverse effects on data quality. Overall, our results suggest that AULP can be an effective tool for increasing contact and response rates, especially for translated surveys, but possible negative socio-cultural attitudes towards lotteries need to be considered.

Keywords: Cross-cultural survey research; Incentivized email recruitment; International survey research; Islamic survey methodology; Randomized controlled trial; Response rate; Contact rate

1 Introduction

The global proliferation of surveys has been particularly pronounced for online surveys, with survey invitation emails currently the fastest growing form of surveys worldwide (ESOMAR, 2018), largely due to: the increasing ubiquity of personal computers and smartphones, in combination with expanded internet access (Jaeckle, Burton, Couper, & Lessof, 2019; Toninelli & Revilla, 2016); and relatively low effort, cost, and time involved on the part of the data-collector (Bonke & Fallesen, 2010; Lazar & Preece, 1999). However, as is well-known in the survey research literature, these factors do not imply success of online surveys in terms of adequate response rates, low non-response bias and sufficient data quality overall. In fact, response rates are low and declining in general (Bonke & Fallesen, 2010; Curtin, Presser, & Singer, 2005; de Leeuw & de Heer, 2002; Hansen, 2006; Zhang, Lonn, & Teasley, 2016), and these problems are particularly acute for online surveys. According to meta-analyses by Daikeler, Bosnjak, and Lozar Manfreda (2019) and Lozar Manfreda, Bosnjak, Berzelak, Haas, and Veho-

Contact information: Robert Lensink, PLEASE PROVIDE POSTAL ADDRESS OF CORRESPONDING AUTHOR (E-mail: b.w.lensink@rug.nl

var (2008), response rates to online surveys are about 11% lower than those to other survey modes. Yet, considering their widespread use by academic researchers and practitioners alike (e.g., online surveys generate about one-third of all market research revenues in recent years, Mazareanu (2019), Morea (2014)), and against the backdrop that alternatives are often infeasible, finding (cost-) effective measures to boost online survey response rates, without compromising data quality, is of paramount importance.

In response, various studies investigate the effects of: (1) incentives (Bonke & Fallesen, 2010); (2) design choices pertaining to survey questionnaires and invitation emails; and (3), how these relate to one another. Research on incentives prominently reflects the fact that a commonly chosen form of incentives for online surveys is lottery incentives (see page 2 of Goeritz and Luthe (2013a), and references therein). This is particularly notable, as evidence suggests that the performance of lotteries is highly context-dependent and meta-studies conclude that their *overall* performance is worse than that of other forms of incentives (e.g. Singer & Ye, 2013). Perhaps not surprisingly then, the choice of lottery incentives for online surveys seems, more often than not, a consequence of constraints such as budgetary and logistical limits that render alternative types of incentives infeasible, particularly in reference to large, global samples (Zhang et al., 2016). Data security concerns, cultural taboos, and legal constraints on other forms of incentives (e.g., upfront ones) also play important roles.

This raises the question of how to improve the effectiveness of lottery incentives for online surveys with invitation emails without adverse effects for data quality. Since subject lines of invitation emails to online surveys play a key role for the important first step of *contact*, a natural and interesting line of exploration for the question then becomes subject line manipulation. This line of enquiry is particularly relevant since the core issue of falling *response rates* for online surveys has been linked to the decline of *contact rates*, due to, among other factors, the growing volume of unsolicited email (Daikeler et al., 2019; Zhang et al., 2016), which has made it progressively harder to capture the attention of potential respondents and induce them to open survey invitation emails (related to the more general issue of so-called email overload).

Our main contribution is to study how contact and response rates, as well as non-response bias and other response quality measures for online surveys, might be affected by announcing undefined lottery prizes (AULP) in the subject line of survey invitation emails. We do so by employing a unique randomized controlled trial involving a global sample of 5,128 key staff members working for microfinance institutions (MFIs).

The decision whether to open a survey invitation email and fill in the associated questionnaire will naturally depend not only on the subject line manipulation alone, but also on the interaction of the subject line manipulation with respondent characteristics, and the characteristics of the survey itself. The idea that contact and response, in an online survey invitation email context, can be heterogeneous due to the above factors is intuitive, and can be understood through the lens of the leverage-salience theory (Groves et al., 2000). According to this theory, a person's decision to participate in a survey is the consequence of a cost–benefit analysis, in which the leverage of any survey attribute reflects the positive or negative view that the potential respondent takes of it, depending on their preferences, and its salience or importance depends on the design features of the survey invitation and how they interact with the person's preferences. On this basis, we expect that the AULP treatment makes the underlying lottery more salient, and should boost contact and response rates on average.

We further form a priori hypotheses about heterogeneity stemming from two areas. First, a novel contribution of our paper is to exploit differences in the cultural perceptions of lotteries among our global sample: specifically, shaped by Islamic religious norms, Muslim socio-cultural attitudes toward lotteries are expected to be more negative. The main idea is that gambling is forbidden in Islam, and the negative perception towards gambling may be partially transferred to lotteries. In fact, there is ample evidence that Muslim socio-cultural attitudes towards lotteries are generally less positive. The implications of this likely difference in perception have been explored in other areas such as in marketing, market research, and finance, but have not been studied in the context of survey design, to the best of our knowledge. Hence, we explore being based in a country with membership in the Organization of Islamic Cooperation (OIC) as a moderating variable, and our hypothesis is that the effect of the lottery-based AULP treatment will be lower for such individuals. In the terminology of leverage-salience theory, individuals whose MFIs operate in a predominantly Islamic environment may assign a less positive leverage to the lottery incentive. Thus, the effect of the salience-inducing AULP treatment on contact and response rates is expected to be lower in states that are members of the OIC, than in non-OIC states. Second, as the provision of translations into a local language can be thought of as enhancing the salience of the underlying lottery further, we hypothesize that it will accentuate both the positive effect of AULP for non-OIC countries, and the reduction in the AULP effect for OIC countries.

Importantly, as a higher response rate by itself does not indicate that a survey was more successful, we investigate possible side effects of the AULP treatment in terms of non-response bias and other measures of data quality (e.g., time variables, filled items ratio, recontact, answer patterns, break-offs, and straightlining).

The remainder of this article proceeds as follows: Section

2 contains a review of relevant literature and formulates hypotheses. In Section 3 we describe the data and outline the statistical methodology. Section 4 provides the main regression results and robustness checks, and summarizes our findings on non-response and other data quality analyses. Section 5 concludes.

2 Background and hypotheses

2.1 Online surveys with invitation emails and lottery incentives

Low and falling response rates are a major concern that *online* surveys, in particular, seem to suffer from (Couper, 2001; Daikeler et al., 2019; Dillman & Bowker, 2001; Petchenik & Watermolen, 2011; Shih & Fan, 2008). In addition, a variety of constraints that researchers and practitioners frequently encounter, make survey invitation emails a widely implemented contact mode for online surveys, despite studies suggesting that mixed contact modes might be preferable (e.g. Kaplowitz, Lupi, & Arreola, 2012). The challenges facing invitation emails as a contact mode have been further compounded by the increased competition for prospective respondents' attention (e.g., due to the overall proliferation of unsolicited emails), as well as "oversurveying" trends evoked by the low cost of online surveying (Couper, 2000; Rogelberg & Stanton, 2007).

To address the issue of low response rates, the survey methodology literature has studied the effect of a great variety of features, taking into account how they may counteract or enhance each other's effectiveness (Dillman, 1978; Dillman, Smyth, & Christian, 2014). These features include: the role of the survey sponsor (e.g. Boulianne, Klofstad, & Basson, 2011), the provision of estimates of the time/effort required (e.g. Kaplowitz et al., 2012), and the mention of deadlines (e.g. S. R. Porter & Whitcomb, 2003). Incentives and financial incentives, in particular, have been shown to be successful (Dillman, Smyth, & Christian, 2009; Singer & Ye, 2013).

The general success of incentives can be thought of as the consequence of respondents performing a cost-benefit calculation embedded in a social fabric where reciprocity and trust matter (Dillman et al., 2014). Incentives work, then, mainly by increasing expected benefits (Church, 1993; M. E. Porter, 2004; Warriner, Goyder, Gjertsen, Hohner, & McSpurren, 1996), and by triggering the norm of reciprocity, which states that people feel obliged to respond to positive actions from another person or feel courteous to respond to gifts (Groves, Cialdini, & Couper, 1992; Groves et al., 2000). The role of trust in this context is determined by the type of incentive considered: for lottery incentives trust is important as it will enhance the expected benefits of participating in a survey; while for prepaid incentives, the trust shown by the researcher can be expected to trigger reciprocity.

Importantly, it has been shown that incentives are able to increase response rates across *all* survey modes (Pforr, 2016; Singer & Ye, 2013), and the literature explicitly suggests their use to increase response rates in online surveys (Church, 1993; Van Horn, Green, & Martinussen, 2009), distinguishing between prepaid, conditional and lottery incentives. Prepaid incentives, especially prepaid *cash* incentives, have been shown to repeatedly outperform other incentives for a variety of survey modes (Baruch & Holtom, 2008; Birnholtz, Horn, Finholt, & Bae, 2004; Church, 1993; Dillman et al., 2009; Kypri & Gallagher, 2003; Parsons & Manierre, 2014; Yammarino, Skinner, & Childers, 1991). However, prepaid incentives are difficult to implement with online surveys (Hoonakker & Carayon, 2013; S. R. Porter & Whitcomb, 2003). Conditional incentives, such that incentives get disbursed after the completion of an online survey, appear to evoke mixed results, according to the relatively scarcer literature studying their effect (Goeritz, 2006; Patrick, Singer, Boyd, Cranford, & McCabe, 2013).

By far the most popular choice of incentive in online survey recruitment is lottery-based incentives, despite the literature suggesting a very mixed and overall not very promising picture (e.g. Goeritz & Luthe, 2013b; Singer & Ye, 2013). In particular, multinational, multicultural online surveys make use of this type of incentive, because other incentive types that are more effective according to the literature, especially upfront ones (Dillman et al., 2014), would come up against logistical and budgetary constraints and even might confront cultural taboos or legal issues, as in our case. Therefore, the literature has studied the conditions and contexts in which lottery incentives in online surveys with invitation emails may be most effective, in comparison to other incentives.

For example, in comparison to upfront incentives, time preferences and trust may matter. Besides, the salience of the incentive may also play a role. As we noted, the leverage-salience theory of survey participation (Groves et al., 2000) provides a useful framework. According to this theory, survey participation decisions result from trade-offs made between positive and negative survey attributes, which are assigned leverage, which refers to how positively or negatively individuals view an attribute depending on their preferences (which can vary across prospective respondents depending on factors such as income, culture, demographics etc.); and further defined by salience, i.e., the importance of the attribute (which is determined by design features of the survey, for example, emphases placed on certain survey attributes by the surveyor, in combination with the before-mentioned respondent attributes). Zhang et al. (2016) posit that unreported variation in the salience levels of lottery incentives and the different leverage assigned to lottery incentives by different sample populations largely might explain the mixed findings in prior studies regarding the impact of lottery incentives on survey participation. To explore the salience aspect, we consider a subject line manipulation, AULP.

2.2 Subject line manipulations

In this subsection we review the literature on subject line manipulation and connect it to our AULP intervention. In line with our discussion of Dillman et al. (2014), a key strategic choice that influences the success of email-based survey recruitment is the careful design of various elements of the invitation email itself. In particular, one critical design feature with potentially large impact on response rates is the email subject line: it is well established in the literature that email recipients use it as a filtering mechanism when deciding whether to open an email (captured by contact), thus determining the upper bound for response (Dillman et al., 2014).Surprisingly, the literature examining the effects of subject line manipulations for survey invitation emails on online survey participation is not as large as one would expect (Sappleton & Lourenco, 2016). Notably, Trouteaud (2004) studied the effect of a "plea" and found that pleas offered a significant increase in response rates. S. R. Porter and Whitcomb (2005) reported the effect of different types of invitation email subject lines (i.e., a plea for help, sponsorship of the e-mail, reasons for the e-mail, and a blank) and found a significantly lower involvement rate for a blank subject line. Sappleton and Lourenco (2016) test the effect of a blank subject line versus a tailored subject line (i.e., placing the research question) but find no effect on overall response rates. Some studies suggest the use of an authority figure for increasing survey response rates (e.g. Joinson & Reips, 2007); still, others find higher response rates with high subject matter salience (e.g. Cook, Heath, & Thompson, 2000; Kaplowitz et al., 2012; Marcus, Bosnjak, Lindner, Pilischenko, & Schuetz, 2007).

Even fewer studies have examined the effect of mentioning (lottery-based) incentive prizes in the subject line of a survey invitation or reminder email, notably Janke (2014) and Zhang et al. (2016). Janke (2014) mainly addresses the effect of naming a *specific* lottery prize, using a sample of university students and faculty in Canada. He conducts a comparison between data from two years in which prizes were mentioned and a prior year in which they were not. Using a measure of valid surveys received, Janke identifies an increase in the overall response rate, between the initial and final two survey rounds, with no evidence of adverse effects on data quality. Yet, Janke (2014) suggests randomized studies as a possible direction of future research by noting in his conclusions that: "*Further research to replicate these findings in other contexts and using an experimental design would be beneficial.*" He does so because he faced several confounding factors in his analysis, including slightly different (in-kind) prizes and descriptions, across comparison groups and years; much shorter surveys in the two latter years relative to the baseline year; and changes to the sampling frame and method across groups and years. Zhang et al. (2016) instead manipulate the presentation of lottery-based incentives with email recipients being part of an incentive-centered treatment (i.e., invitation text focuses on the prize) or a survey-centered treatment (i.e., text focuses on aspects such as the importance of respondents' feedback for the community). They find significantly higher response rates to the incentive-centered treatment, especially among low-income respondents, and only minor effects on data quality. However, their experiment changes the survey invitation email text, as well as the subject line, between treatments. The sample they use includes employees of the university institution sponsoring the survey, and the incentive provided is a gift certificate.

To build on and extend this literature, we focus instead on a single, specific variation: manipulating the subject line of survey invitation emails by announcing undefined lottery prizes (AULP), compared with a subject line that is identical in all respects except that it does *not* announce the undefined lottery prizes. Our treatment is designed, in terms of the leverage-salience theory introduced earlier, to increase the salience of the underlying lottery, and our particular approach of mentioning an unspecified prize in the subject line is inspired by the information gap theory of cognitive processing by individuals, by which lack of information can trigger curiosity (S. R. Porter & Whitcomb, 2003; Sappleton & Lourenco, 2016), in our case among potential survey respondents, and prompt them to open the invitation email. Notably, AULP was implemented in the initial survey invitation email, as well as in up to four reminders that were sent in case individuals did not fill or incompletely filled in the survey (as follow-up contact has been suggested to increase survey participation; Van Mol (2017), Yammarino et al. (1991)).

Thus, though we address some related topics, important differences distinguish our paper from Janke (2014) and Zhang et al. (2016), pertaining to the research question, sample, design, analysis, methodology, and results. In particular, we mention the possibility of winning an unspecified prize in our email invitation subject line and use a non-student, global sample (cf. Janke, 2014), which enables us to analyze the implications of cultural heterogeneity when it comes to the leverage people assign to lotteries (see Subsection 2.3). We also remedy a key shortcoming that Janke (2014) mentions, by providing experimental evidence based on the randomized AULP treatment. Furthermore, whereas we explore leverage and salience similar to Zhang et al. (2016), our treatments and samples differ from theirs in relevant ways, and we use workplace-related incentives.

2.3 Islamic religious norms toward gambling and perception of lotteries

Our global sample allows us to add a distinctive and novel element to the investigation of the AULP effect, which is the study of the implications of attitudinal differences toward lotteries across cultures. According to leverage-salience theory, prospective respondents with different preferences, as they tend to be shaped by culture, socio-economic characteristics, and demographics, etc., might assign different leverage to the same survey attribute, such as a lottery for prizes. It is therefore useful to consider the role of such moderating variables. Some likely moderators include age (i.e., the relevance of a survey topic may vary by age) and sense of civic duty (i.e., the degree to which someone is motivated by incentives may vary with their civic sense, from Groves et al. (2000). Other examples from the literature, specifically to do with lottery prizes, are income (i.e., relatively low-income individuals may assign higher leverage to incentives), from Zhang et al. (2016), and gender (i.e., in the case of online surveys, females appear to respond more to lottery incentives), from Heerwegh (2006) and Laguilles, Williams, and Saunders (2011).

Turning to the specific case of lotteries, there is well established evidence that Muslim socio-cultural attitudes towards lotteries are generally less positive, as gambling is forbidden in Islam, and the negative perception towards gambling may be partially transferred to lotteries. As Hassanat and Al tarawneh (2015) explain, "According to Islamic values, Muslims normally do not play traditional lottery, because it is considered as gambling, which is taboo in Islam." Notably, while gambling is taboo in a number of religions besides Islam, there are important qualitative differences: explicit legal restrictions against gambling; often harsher and more discouraging societal attitudes towards gambling; and considerable restrictions (and often stigma) associated with visits to venues such as casinos (which, if they exist in OIC member states, primarily cater to tourists and non-Muslim locals, see Hassanat and Al tarawneh (2015). Thus, on average, respondents whose MFI operates in a predominantly Islamic environment are expected to assign a different leverage to lotteries than respondents whose MFI does not operate in a predominantly Islamic environment.

For this study, we therefore explore whether the location of a prospective respondent's MFI, in an OIC member state or not, moderates the treatment effect of the salience-inducing AULP treatment (see Appendices A.1 and A.2 for the list OIC member and non-member states). We anticipate that the effect of AULP is weaker in OIC member states than in non-member states (without forming an a priori expectation whether the overall AULP effect for OIC countries is positive or negative). To the best of our knowledge, we are the first to propose this effect. While the importance of accounting for Muslim socio-cultural attitudes and Islamic religious norms is widely recognized in other research fields, such as marketing and market research (e.g. Young, 2007) or finance (Ahmad, Lensink, & Mueller, 2020), their effect on the perception of lotteries as incentive mechanisms appears to have been neglected in the literature on the use of lottery incentives to boost survey participation. This is especially surprising given the popularity of lottery incentives for online survey recruitment, where multicultural and multinational surveys are common. Thus, our line of investigation is of critical importance as it might highlight a drawback of a class of prize assignment mechanisms that has become predominant for a wide range of contexts involving participant recruitment—such as (online) surveys or crowdsourcing—in which invitation emails are distributed to large pools of potential participants.

2.4 Translations

In comparative survey design studies, translations constitute a key success element for multinational and multicultural surveys, because language barriers are a critical impediment to survey completion (Fowler, 2012; Harkness, van de Vijver, & Mohler, 2002; Harkness, Villar, & Edwards, 2010). We provided translations in a national/official language of the state in which potential respondents' MFI was located, together with the English default version, for a subset of the 124 countries represented in our sample (see Subsection 3.2, Appendix A.3, and Table A1). This allows us to consider translation as a moderating variable,[1] with the prediction that translations increase the salience of AULP which by itself increases salience of the underlying lottery. Therefore, translations should accentuate both the positive effect of AULP among non-OIC countries and the reduction of the AULP effect for OIC countries. This analysis provides an additional interesting layer to our study of subject line manipulations in the context of online surveys with survey invitation emails and lottery incentives.

2.5 Data quality

With exploratory analyses, we investigate whether the AULP treatment produces any unwanted side effects in terms of data quality (e.g., leverage effects that we did not foresee). We explore this possibility in terms of non-response bias (except for OIC countries, systematic non-response patterns for which is one of our main hypotheses as explained in Subsection 2.3) as well as several other measures of data quality, such as missing items and speeding.

2.6 Hypotheses

On the basis of the preceding discussions in Subsections 2.1-2.4, we form three main hypotheses. First, due to increased salience, AULP increases contact and response rates

[1] According to existing literature (e.g. Harzing, 2000), reducing language barriers by providing translations is in itself expected to boost survey participation. However, as we do not randomize the provision of translations for our study (see Subsection 3.2), we do not investigate this question.

overall (Hypothesis 1). Second, since AULP is based on the offering of lottery incentives, OIC membership is expected to play a role in the perception of AULP due to cultural factors. Thus, the effect of AULP on contact and response rates is attenuated in the OIC subgroup (Hypothesis 2). Third, translations should induce further salience, which may moderate these AULP effects. In particular: the AULP treatment effect on both contact and response rates is greater among the translated, non-OIC subgroup (Hypothesis 3a); and, the reduction of the AULP effect among OIC countries is more pronounced when translations are provided, driven by increased salience due to translation of AULP (Hypothesis 3b).

3 Data and methodology

3.1 Sample selection

The sample selection relied on MIX Market (https://www.themix.org/mixmarket), an online microfinance information platform that provides data about a comprehensive sample of MFIs from all over the world. Its existing procedures ensure data quality (through a system of ratings) and comparability across countries.[2] We purchased a list of all 2,641 MFIs reporting to MIX Market, along with contact details of 5,649 key staff members (often CEOs and other top managers with diverse job titles) of these institutions. After dropping rows with incomplete, missing, or duplicate email addresses, the sample contained 5,128 unique email addresses of microfinance practitioners, representing 2,527 MFIs located in 124 states (see Appendices A.4, A.5 and Figure A1). While these key decision makers are somewhat comparable to top managers of commercial financial institutions, they feature some notable differences, too. In particular, on average, the CEOs, other executives and other practitioners of MFIs tend to have more of a social entrepreneurship or philanthropic background, rather than a banking background that is more common among CEOs of commercial banks. The percentage of female CEOs is higher among MFIs, and a relatively high percentage of them also are board presidents (Beisland, Ndaki, & Mersland, 2019; Strom, D'Espallier, & Mersland, 2014). We discuss the characteristics of MFI practitioners and the implications of these for the wider applicability of our results in Section 5.

3.2 Translations

The subject lines, emails themselves, and online surveys all were written exclusively in English or in both English as well as an official language (other than English) of the recipient MFI's state. The provision of translations was not randomized, but largely reflected budgetary constraints. We describe the translation criteria in great detail in Appendix A.3. In each case in which we provided a translation, the English version appeared second. We classified all 5,128 unique email addresses into 10 language groups (see Appendices A.1, A.2 and Figure A1): English (n = 1,790) and 9 non-English language groups (e.g., Arabic, French, etc.; n=3,338) according to the criteria in Appendix A.3. All translations were conducted by official translation agencies and verified with back-translations. Any inconsistencies were corrected using a reconciliation process involving the translators, back-translators, and members of the author team.[3] The translations cost approximately €1,000 in total.

3.3 Survey logistics

All 5,128 email addresses were sent a personalized invitation email, which included a link to an online survey. The source of the invitation email was an easily verifiable, official work email address, making it unlikely to be considered spam. To increase confidence in, and the perceived credibility of, the study, the email signature included an official emblem of the sender's workplace, as well as contact information for all members of the research team. Official sponsorship of this kind has been shown in the literature to increase survey participation in both mail and online surveys (Fox, Crask, & Kim, 1988; Goyder, 1982; Heberlein & Baumgartner, 1978; Lozar Manfreda et al., 2008; Walston, Lissitz, & Rudner, 2006). All the information provided to respondents in the context of this study was accurate and truthful; we did not use any deception.

AULP treatment: subject line manipulation of the survey invitation email. The sole difference between the treatment and control groups was the subject line for the invitation email. In the AULP treatment group, it read "5-minute Survey on MFIs—Exciting Prizes." The subject line for the control group instead read "5-minute Survey on MFIs." Therefore, both subject lines offered an estimate of the time needed to complete the survey. The short questionnaire reflects the principles of social exchange theory, which posits that prospective respondents are more likely to participate in a survey if they think the survey is short (Marcus et al., 2007). Moreover, the subject line conformed to the principle of "topic salience" in both cases, which also should enhance the response rate, because people are more likely to respond when the survey topic is of importance to them (Martin, 1994; Sheehan & McMillan, 1999; Walsh, Kiesler, Sproull, & Hesse, 1992).

[2]The MIX Market also has the greatest coverage (cf. alternative sources such as CGAP, CGAP's Financial Development Gateway, Microcredit Summit, or SEEP Network) of three key features of MFIs: decision makers, financial indicators, and social performance indicators.

[3]Back–translations remain a standard tool for assessing the quality of international/cross-cultural surveys (Tyupa, 2011), despite some recent critiques (Harkness et al., 2010). Subject to logistical and budgetary constraints, we take these critiques partially into account, by adopting a team-based reconciliation process.

Main text of the survey invitation email. The main text of the survey invitation email asked recipients to respond to our survey and provided information about the following work-related, non-monetary as well as monetary, prizes relevant for our sample that could be won via lottery:

1. One first prize: the right to attend an international summer school on microfinance at a research university ranked in the top 100 worldwide, including all travel and lodging expenditures.
2. One second prize: identical to the first prize, but without travel expenditures.
3. Six third prizes: free access to instruction materials used for summer school.
4. Ten fourth prizes: a donation of €50 to each winning MFI in this prize category.

A web link embedded in the email offered further prize details. The respondents also had the option to fill in the survey but opt out of entry into the lottery. The complete text of the email is in Appendix A.6 (Figure A2); we provide a detailed breakdown of the expected and actual costs of the incentives, along with information about prize consumption, in Appendix A.7.

Reminder emails. The first round of invitation emails was sent on 6 April 2016. Up to four follow-up reminders were sent at intervals of 12 days, each with the same subject line as the original invitation email. Sending these reminders was a costless feature of the survey platform we used (Qualtrics). Although these effects are not the main focus of our analysis, we find that the four follow-up reminders increased the response rate by 85% compared with the initial email (Appendix C, Tables C1 and C2). Importantly for data quality, Qualtrics tracked the identity of respondents and sent reminders only if they had not already completed the survey. Furthermore, respondents could pause and resume the survey at will, with no loss of prior responses. These features eliminate the risk of duplicated responses due to the reminders. Finally, in each round, recipients could unsubscribe from the survey mailing list.

3.4 Randomization into treatment (AULP) and control groups

We randomly assigned half of the 5,128 email addresses to the AULP treatment group and half to the control group, stratifying by language groups. Each potential participant remained in the same treatment group throughout the study.

3.5 AULP treatment–control balance

We collected data from the MIX Market database pertaining to key characteristics of potential respondents, participating MFIs, and states in which those MFIs were located, at the time of the study. We used these variables to check whether the allocation of potential respondents into the AULP treatment and the non-AULP control group was random. Table A2 in the Appendix presents these balancing tests, i.e., for each pertinent variable, a statistical test for the differences in means between treatment and control groups (along with their overall means; means for the treatment and control groups; and standard deviations). As the results show, the randomization was successful: Of the 19 variables tested, we find an imbalance for just a single variable at a 5 percent significance level (*Mature*, a dummy variable equal to 1 if the MFI is older than 8 years, and 0 otherwise) and another at the 10 percent significance level (*Target Market*, a dummy variable equal to 1 if the target market of an MFI is *broad/high end*, and 0 otherwise). At a 90 percent confidence level, on average 2 out of every 20 variables will be unbalanced by statistical chance, so our randomization appears quite successful. Nonetheless, to be cautious, we include the two unbalanced variables as controls in our regression analyses.

3.6 Main outcome rates and variables

The two main outcomes of interest are contact and response. As a subject line manipulation, AULP should directly affect the decision whether to open the survey invitation email (captured by the contact rate). Conditional on opening it, for a given respondent, survey completion depends on a multitude of factors, with contact rates being greater than or equal to response rates by definition.

To calculate the contact rate, we use AAPOR's 2016 standard definition of Contact Rate 1 (i.e., the minimum contact rate without eligibility adjustment), the most conservative of all the possible AAPOR definitions. That is, Contact Rate 1 counts everyone who opened the invitation email in the numerator and counts all emails sent in the denominator, i.e., it does not exclude category 3.30 ("invitation returned undelivered") or allow for any (subjective) eligibility adjustments to category 3.19, "nothing ever returned" (e.g., excluding an estimated number of emails that went to spam folders). For our regression analysis, the binary variable "Contact" thus is defined over all emails sent, and it takes a value of 1 for everyone who opened the email. We calculated the response rate in line with AAPOR's 2016 standard definition of Response Rate 1 (i.e., the minimum response rate without eligibility adjustment), which like our contact rate, is the most conservative possible, since it only counts fully completed surveys in the numerator, and the denominator is the same as for the contact rate. In the regression analysis, the binary variable "Response" is defined accordingly over all emails sent; equal to 1 only for fully completed surveys. For some of the corroborating analyses, we use Cooperation Rate 1, calculated according to AAPOR's 2016 standard definition as the ratio of fully completed surveys to all opened emails. Further details about how we calculated these three rates are available in Appendix B.1. Table 1 tabulates the data needed to calculate our contact and response rates. The overall Contact Rate 1 is 43 percent (Appendix B.1, Table B1), notably higher compared to other recent email-based studies in the financial sector, such as the Cloud (2016) report that finds an overall contact rate of only 22 percent.

In line with AAPOR guidance, we define break-offs, partial, and complete cases using the following criteria: If less than 50% of all essential or crucial questions are answered (excluding a refusal or no answer), it constitutes a break-off; 50–99% answered constitutes a partial; and 100% answered is complete. As Table 1 shows, 606 surveys are complete, 94 are partials, and 92 represent break-offs. In turn, we determine an overall Response Rate 1 of 12 percent. The definition of Response Rate 1 should be borne in mind when judging this low number; it may partially reflect the conservativeness of Response Rate 1.[4]

Furthermore, the intervention only pertains to the subject line manipulation of the survey invitation email (which can be expected to influence contact); but the likelihood of fully completing the survey (i.e., of response) is additionally influenced by many other factors, e.g., characteristics of the main body of the invitation email, the prizes offered, as well as the survey itself, including what information is asked for etc. For example, our survey was directed towards key decision-makers in MFIs (who presumably have high opportunity cost for their time) and asked for financial and operational information about their firms (which the respondents presumably would need to expend some mental effort in replying to). These factors might have contributed to the low response rate.

The cooperation rate for our survey, calculated according to AAPOR's standard definition of Cooperation Rate 1, is 28 percent (Appendix B.1).

3.7 Main explanatory and control variables

Our main explanatory variable of interest is the randomized AULP treatment indicator (=1 for the treated group, 0 otherwise). In addition, we analyze the moderation of the AULP treatment effect for staff members of MFIs operating in OIC member states (=1, 0 for non-member states) and the provision of translations in an official language other than English (=1, 0 if no translation is provided). In all our regressions, we include the two variables that were unbalanced in the balancing table as controls (*Mature* and *Target Market*; see Subsection 3.5). Other variables used in our analyses are described in the corresponding sections: for sensitivity see Subsection 4.2; for non-response and other data quality analyses see Subsection 3.8, and more in detail, Appendix D.

3.8 Data analysis methodology

We analyze the data from our experiment using the following steps.

Main regression analyses. To formally test Hypotheses 1–3, as listed in Subsection 2.6, we estimate three logistic regression models using the full sample of individuals to whom the invitation email was sent (n = 5,128, Table 1) and the two binary dependent variables, "Contact" and "Response," as defined in Subsection 3.7. For each individual that the invitation email was sent to, denote the probabilities that the dependent variables "Contact" and "Response" take values of 1 (i.e., the likelihoods of contact and response), by p_C and p_R, respectively. The logistic regressions model the log-odds, $\log(\frac{p}{1-p})$, where p can be either p_C or p_R, as a linear function, $x'\beta$, of a vector of covariates x. The regressions described hereafter all include the two unbalanced variables as controls (*Mature* and *Target Market*; Subsection 3.5). Because we are not interested in the coefficients of these two control variables, we describe the relevant terms in Models 1–7 below as "Controls" for simplicity. The first logistic model,

$$\log\left(\frac{p}{1-p}\right) = \beta_{0.1} + \beta_{1,1}\text{AULP} + \text{Controls} + \varepsilon_1 \quad \text{(Model 1)},$$

tests Hypothesis 1, or our prediction that AULP has positive average treatment effects on both contact and response, or that $\beta_{1,1} > 0$.

To determine whether the AULP treatment effect is moderated by cultural factors, such that it is lower in the OIC member state subgroup than in the non-OIC member state subgroup (Hypothesis 2), we estimate another logistic model:

$$\log\left(\frac{p}{1-p}\right) = \beta_{0.2} + \beta_{1.2}\text{AULP} + \beta_{2.2}\text{OIC} + \beta_{3.2}\text{AULP} \cdot \text{OIC} + \text{Controls} + \varepsilon_2 \quad \text{(Model 2)}.$$

In this case, as a statistical hypothesis, we predict that $\beta_{3.2} < 0$. The AULP treatment effect in the non-OIC subgroup is given by the coefficient $\beta_{1.2}$, whereas in the OIC subgroup, it is given by $\beta_{1.2} + \beta_{3.2}$. If their difference is negative, $\beta_{3.2} < 0$, it would indicate that the AULP treatment effect is *smaller* in magnitude in the OIC subgroup than in the non-OIC subgroup.

Finally, we explore if the AULP treatment effect in non-OIC countries and the moderating influence of OIC membership on the AULP treatment effect vary across translated and non-translated subgroups. To do so, we test if the AULP treatment effect is differently moderated in the subgroups of

[4] For example, counting category 1.2 ("partial or break-off with sufficient information") in the numerator and not counting emails that bounced back (category 3.30, "invitation returned undelivered") in the denominator, when calculating the response rate, would have led to a response rate of 16 percent. The response rate also would have increased substantially for even a very modest eligibility adjustment to category 3 ("nothing ever returned"), because 2,054 observations fall into this category.

Table 1
Outcome rates

Survey Status (AAPOR 2016 Disposition code)	N	%
Returned Questionnaire (1.0)		
Complete (1.1)	606	12
Partial or break-off with sufficient information (1.2)	94	2
Eligible, "Non-Interview" (2.0)		
Refusal (2.11)		
Explicit refusal (2.111)	40	1
Implicit refusal (2.112)		
Logged on to survey, did not complete any items (2.1121)	141	3
Read receipt confirmation, refusal (2.1122)	1217	24
Break-off or partial with insufficient information (2.12)	92	2
Unknown Eligibility, "Non-Interview" (3.0)		
Nothing ever returned (3.19)	2054	40
Invitation returned undelivered (3.30)	884	17
Total	5128	100

This table shows the breakdown of the sample (numbers per group and percentages of the total sample per group) according to AAPOR's 2016 disposition codes. The unit of observation is an MFI practitioner. Data source: Authors' survey.

data defined jointly by the OIC and Transl variables, using the following logistic model:

$$\log\left(\frac{p}{1-p}\right) = \beta_{0.3} + \beta_{1.3}\text{AULP} + \beta_{2.3}\text{OIC} + \beta_{3.3}\text{Transl} + \beta_{4.3}\text{AULP}\cdot\text{OIC} + \beta_{5.3}\text{AULP}\cdot\text{Transl} + \beta_{6.3}\text{Transl}\cdot\text{OIC} + \beta_{7.3}\text{AULP}\cdot\text{Transl}\cdot\text{OIC} + \text{Controls} + \varepsilon_3 \quad \text{(Model 3)},$$

where the left-hand side of the regression is the same as for Models 1 and 2. Our two statistical hypotheses are $\beta_{5.3} > 0$ (Hypotheses 3a) and $\beta_{7.3} < 0$ (Hypotheses 3b). That is, Hypothesis 3a predicts that the AULP treatment effect on both dependent variables (contact and response rates) is higher in the *translated* (cf. non-translated) non-OIC subgroup. The AULP treatment effect in the non-translated non-OIC subgroup is given by $\beta_{1.3}$, whereas in the translated non-OIC subgroup, it is given by $\beta_{1.3} + \beta_{5.3}$. If their difference were positive, $\beta_{5.3} > 0$, it would imply, for the non-OIC subgroup, that the AULP treatment effect is greater in magnitude when translations are provided compared to when no translations are provided. Hypothesis 3b predicts that the reduction of the AULP effect we hypothesize in OIC countries (Hypothesis 2) gets accentuated when translations are provided (i.e., $\beta_{7.3} < 0$), driven by the increased salience due to the translation.

Additional regression analyses. We estimate a series of additional regressions to perform sensitivity analyses of the main regression analyses, as well as check for possible effects on data quality. To confirm if the main results regarding the OIC variable are robust, we control for two cultural characteristic scores, Individualism and Trust (explained in more detail in Appendix B.2). For this analysis, we use multi-level modeling, because the culture scores are defined at the country level. In particular, Model 4 is the multi-level estimate of Model 3, but we label it separately as Model 4 to avoid confusion. We do not write it out again; it is identical to Model 3 in all other respects. Then we augment Model 4 with the two cultural characteristic variables, alone and in interaction with the AULP treatment, as follows:

$$\log\left(\frac{p}{1-p}\right) = \beta_{0.5} + \beta_{1.5}\text{AULP} + \beta_{2.5}\text{OIC} + \beta_{3.5}\text{Transl} + \beta_{4.5}\text{Individualism} + \beta_{5.5}\text{Trust} + \beta_{6.5}\text{AULP}\cdot \mathit{OIC} + \beta_{7.5}\text{AULP}\cdot\text{Transl} + \beta_{8.5}\text{Transl}\cdot\text{OIC} + \beta_{9.5}\text{AULP}\cdot\text{Transl}\cdot\text{OIC} + \beta_{10.5}\text{AULP}\cdot\text{Individualism} + \beta_{11.5}\text{AULP}\cdot\text{Trust} + \text{Controls} + \varepsilon_5 \quad \text{(Model 5)}.$$

As an important check of data quality, we explore potential non-response bias, by estimating a logistic model for the log-odds ratio of non-response as follows:

$$\log\left(\frac{p_{NR}}{1-p_{NR}}\right) = \beta_{0.6} + \beta_{1.6}\text{AULP} + \beta_{2.6}X + \beta_{3.6}\text{AULP}\cdot X + \text{Controls} + \varepsilon_6 \quad \text{(Model 6)}.$$

Here, p_{NR} is the likelihood of non-response, or the probability that a binary variable denoting non-response, NR, takes a value of 1. It equals 1 for category 2.0 (eligible, "non-interview") and 0 for category 1.0 (returned questionnaire), according to AAPOR 2016 disposition codes. Thus, NR indicates whether a particular email recipient was a non-respondent or respondent. As a robustness check, we also comment on the results of an alternative specification of non-response, where NR takes a value of 1 for categories 2.0 (eligible, "non-interview") and 3.0 (unknown eligibility, "Non-Interview") and 0 for category 1.0 (returned questionnaire), according to AAPOR 2016 disposition codes. Non-response bias might occur if potential participants who did not respond (for whom NR = 1) are systematically different from those who participated (for whom NR = 0). Using the preceding logistic model, we thus check whether an email recipient with a particular observable characteristic X is significantly more or less likely to not respond (e.g., Groves, 2006). In the non-AULP group, we can check whether the coefficient $\beta_{2.6}$ is statistically significant. For the AULP group, we also test if the AULP treatment itself affects non-response bias, by addressing whether the interaction term $\beta_{3.6}$ between the AULP treatment and a particular observable characteristic X is statistically significant. The variables X from the MIX Market database are the ones we used in the balance test (Table A2 in the Appendix). We run regressions for each X separately and present the results in the Appendix.

Finally, we perform additional checks of data quality by estimating an ordinary least squares (OLS) model separately for several non-binary measures of data quality, denoted by "Dquality":

$$\begin{aligned}\text{Dquality} = {} & \beta_{0.7} + \beta_{1.7}\text{AULP} + \beta_{2.7}\text{OIC} + \beta_{3.7}\text{Transl} \\ & + \beta_{4.7}\text{AULP} \cdot \text{OIC} + \beta_{5.7}\text{AULP} \cdot \text{Transl} \\ & + \beta_{6.7}\text{Transl} \cdot \text{OIC} + \beta_{7.7}\text{AULP} \cdot \text{Transl} \cdot \text{OIC} \\ & + \text{Controls} + \varepsilon_7 \quad \text{(Model 7)}.\end{aligned}$$

With this analysis, we can check if any of the coefficients associated with the AULP treatment are statistically significant, with a sign suggesting lower data quality, as well as identify *any* clear patterns of differences in the data quality measures for the treatment versus the control group, as well as for the moderating variables. The Dquality variables, as motivated in Appendix D.3, are as follows: log of time spent completing the survey (LnTime), log of time spent completing the survey divided by the total number of items ($\frac{\text{LnTime}}{\text{Total Items}}$) to account for skip patterns, the winsorized log of time spent completing the survey (WinLnTime), and the winsorized log of time spent completing the survey divided by the total number of items ($\frac{\text{WinLnTime}}{\text{Total Items}}$) to account for time outliers (Appendix Table D5), the filled items ratio, or the ratio of answered to total number of questions/items in the survey (*Filled items ratio*), and the stated willingness of respondents to be contacted again (Recontact). For this last, binary variable, we use a logistic model, in which the left–hand side of Model 7 above is replaced by $\log\left(\frac{p_{\text{recon}}}{1-p_{\text{recon}}}\right)$, where p_{recon} is the probability that the binary variable denoting the stated willingness of respondents to be contacted again takes a value of 1. We present the results of all data quality analyses, including estimates of Models 6 and 7, in the Appendix D.

Taking correlation within MFIs into account. Finally, to account for the possible correlation of the error terms within MFIs, we cluster the standard errors for all our regressions at the MFI level, except for in the multi-level model.

4 Results of regression analyses

We first present the results of the formal tests of Hypotheses 1–3 (Subsection 2.6) using Models 1–3. We first test for treatment effects of AULP and quantify these effects using the empirical estimates of Model 1. Then, we ask, using the estimates of Model 2, if the AULP treatment effect is moderated by cultural factors, in particular, if the treatment effect is lower in the OIC member state subgroup compared to the non-OIC subgroup. Next, we explore using the estimates of Model 3 whether the AULP treatment effect is moderated in subgroups defined jointly by OIC and translation. Thereafter, we perform sensitivity tests of the main results using the estimates of Model 4 (multi-level version of Model 3) and Model 5. We also summarize our main findings regarding Models 6 and 7, as well as some findings about other data quality measures.

4.1 Testing our three main hypotheses

In Table 2, we provide the estimation results for our main logistic regression models, which we use to test Hypotheses 1–3. For ease of interpretation, we report the marginal effect for each independent variable (holding the other variables at their means).

The Model 1 results confirm Hypothesis 1. As Columns (1) and (2) in Table 2 show, the AULP treatment has a positive, statistically significant effect on contact and response rates. This suggests that the AULP treatment makes the lottery more salient, and thereby improves contact and response, raising the contact rate by 2 percentage points above the control group (significant at 1%) and the response rate by 2 percentage points above the control group (significant at 5%). For the significance calculations, we account for possible correlation of the error term within MFIs by clustering standard errors at the MFI level.

As noted in Subsection 3.6, the absolute values of response rates in our sample are low for several reasons that we listed. To get a sense of the economic significance of these findings, it is useful to compare the 2 percentage point change to the response rate in the control (non-AULP) group:

Table 2
Testing main hypotheses

		Model 1 Direct treatment effects		Model 2 Moderation due to culture		Model 3 Additional moderation due to translation	
Variable	n[g]	(1) Contact[e]	(2) Response[f]	(3) Contact[e]	(4) Response[f]	(5) Contact[e]	(6) Response[f]
AULP[a]	2564	0.094***	0.019**	0.116***	0.023*	0.060**	0.028
		(0.014)	(0.009)	(0.018)	(0.012)	(0.027)	(0.018)
OIC[c]	1780	-	-	0.117***	0.038***	−0.025	−0.009
		-	-	(0.022)	(0.014)	(0.046)	(0.032)
AULP×OIC	887	-	-	−0.058**	−0.008	0.038	−0.021
		-	-	(0.029)	(0.018)	(0.060)	(0.041)
Transl[b]	3338	-	-	-	-	−0.018	−0.005
		-	-	-	-	(0.026)	(0.018)
AULP×Transl	1669	-	-	-	-	0.096***	−0.010
		-	-	-	-	(0.036)	(0.024)
Transl×OIC	1390	-	-	-	-	0.183***	0.059*
		-	-	-	-	(0.052)	(0.036)
AULP×Transl×OIC	692	-	-	-	-	−0.145**	0.019
		-	-	-	-	(0.070)	(0.046)
Controls[d]		Yes	Yes	Yes	Yes	Yes	Yes
Observations		5128	5128	5128	5128	5128	5128
McFadden R2		0.008	0.003	0.014	0.007	0.019	0.010

Marginal effects computed at the sample means from three logistic regression models (Models 1–3) with standard errors clustered at the MFI level in parentheses. See Table B1 in the Appendix for data sources.
[a] Dummy variable for potential respondent is (randomly) provided with an announcement of undefined lottery prizes in the subject line of the survey invitation emails.
[b] Dummy variable for potential respondent received a survey translated into an official language of their country and 0 otherwise (see Appendix A.3 for criteria). [c] Dummy variable for potential respondent is a staff member at an MFI is operating in an OIC member state. [d] Dummy variables for target market of an MFI is broad/high end and for MFI is older than 8 years. [e] Dummy variable for potential respondent opened the survey invitation email.
[f] Dummy variable for potential respondent fully completed the survey. [g] Number of observations coded as 1 on the independent variable. For interactions: number coded as 1 on all variables that constitute the interaction.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

The AULP treatment raises response rates by approximately 18 percent compared with the mean response rate in the control group (which equals 11 percent; calculation not shown for brevity). Similarly, the contact rate increases by approximately 25 percent relative to the mean contact rate in the control group (37 percent; calculation not shown for brevity).

For Model 2, our focus is on the sign and magnitude of the marginal effect of the interaction of AULP and OIC. As we explained in Subsection 3.8, a negative sign means that the AULP treatment effect is lower in the OIC subgroup than in the non-OIC subgroup. Column (3) of Table 2 shows that the marginal effect of the interaction term is indeed negative and statistically significant at the 5 percent level for contact rates, which implies that the AULP treatment has a weaker effect on contact rates in the OIC subgroup than in the non-OIC subgroup. We attribute this lower treatment effect to cultural differences between OIC and non-OIC states, such as possible negative associations with lottery-style prize allocation mechanisms that might be perceived as similar to gambling in OIC states. In other words, on average, lotteries are assigned a more negative leverage by people in OIC states. Of course, this result can also be interpreted as the AULP treatment exerting a stronger effect on contact rates in the non-OIC subgroup than in the OIC subgroup. Furthermore, as Column (4) reveals, the marginal effect is not statistically significantly different from 0 for the response rate. That is, the AULP treatment effect on the response rate is not statistically different between OIC member and non-member states.

In Model 3, we find evidence of heterogeneity in the AULP treatment effects, in line with our Hypotheses 3a and 3b. Columns (5) and (6) show that the treatment effect of AULP is moderated in the translated subgroups for both OIC and non-OIC countries. In non-OIC countries, the treatment effect on contact rates from Model 2 increases in the translated subgroup, according to the positive (statistically significant at 1%) marginal effect of the interaction term between AULP and Transl. We thus find evidence in support of Hypothesis 3a. Similarly, the effect identified in Model 2 is accentuated for the translated subgroup in OIC countries, according to the large, negative (statistically significant at 5%), marginal effect of the triple interaction term among AULP, Transl, and OIC. This evidence supports our Hypothesis 3b. In fact, the marginal effect of the interaction between AULP and OIC is not statistically significant, which means OIC countries do not have a significantly different AULP treatment effect than non-OIC countries when translations are not provided. This shows that the lower treatment effect for the OIC countries in Model 2 is in fact driven by the *translated* OIC subgroup. This suggests that translations make the lottery incentive more salient in both OIC and non-OIC countries, but are detrimental in the case of OIC countries, due to the more negative leverage attached to lotteries there. We next report on several robustness checks for this key finding.[5]

4.2 Sensitivity analyses using culture scores

A concern related to the preceding analysis is that the OIC variable—with its key role in our paper—might proxy for cultural differences other than Islamic norms. In particular, cultural characteristics such as individualism–collectivism and social trust might be confounded with the OIC variable. To check the sensitivity of our results to this possible concern, we explicitly control for the two alternative measures of cultural characteristics in Model 5 (as described in Subsection 3.8).

The two culture scores we use were constructed by Beugelsdijk and Welzel (2018), using data from the European Value Studies and World Values Surveys. *Individualism* indicates the degree to which people in a country consider themselves autonomous personalities (higher score) or members of close-knit communities (lower score). *Trust* measures the degree to which they feel comfortable (higher score) or stressed and anxious (lower score) in unstructured situations. Both indices are measured on scales from 0 to 100, and the individual country scores are in Table B2 in Appendix B.2 (where we also describe and motivate the use of the two indices in more detail). Because the scores are available only at the country level, all participants within a single country are assigned a single value for this score.[6] Noting the nested nature of the data used in Model 5, we estimate it with multi-level modeling. As our individual-level outcome variables (contact, response) are binary, we fit a logistic model. Then, we are able to model the outcome probabilities as functions of unobserved and observed individual-level (e.g., AULP) and country-level (e.g., Trust) variables, with random intercepts at the country level.

The results in Table 3 (Columns (3–8)), presented as Models 5a, 5b, and 5c, reveal the outcomes when *Individualism, Trust,* and both are added to interactions with AULP (see Subsection 3.8). For comparison, the results of Model 4 (i.e., Model 3 estimated using multi-level modeling), are presented in Columns (1) and (2), estimated for the same sample as Models 1–3 from Table 2. The sample sizes for Models 5a–5c are much smaller, and they vary slightly across these models due to the availability of the cultural dimension scores, as noted above. The key finding to be taken from Table 3 is that the results for Models 5a–5c are consistent with the results of Model 4. That is, the results, especially for the relevant country-level variables, remain robust to controlling for culture scores. The lone exception is AULP, which becomes insignificant in some of the specifications in Models 5a–5c. The insignificance of AULP can be attributed to the sharp drop in sample size due to missing data for the culture scores in the case of several countries, as noted. However, note that the sign and statistical significance of the triple interaction terms among AULP, Transl, and OIC are consistent between Model 4 and Models 5a–5c. The evidence in favor of Hypothesis 3b thus is robust to the inclusion of the culture scores.

To further take unobserved country effects into account, we replicate all of Table 2 with multi-level modeling. The results remain unchanged (see Appendix Table C3). We also estimate Model 3 from Table 2 using country fixed effects as an alternative to the multi-level model. This model controls for all unobserved time-invariant aspects of the countries in the sample. The results pertaining to the interaction terms remain unchanged (OIC and Transl only vary across countries, so their own effects cannot be estimated using the country fixed effects model).[7]

4.3 Summary of data quality analyses

We perform additional analyses to check whether our treatment, AULP, leads to unwanted side effects in terms of data quality, using non-response bias and several other measures. Our main results can be summarized as follows.

[5]Our models in Table 2 have low McFadden R2 values. This implies that our models are not doing a good job in terms of predicting our dependent variables, which is quite often the case for regression analyses in experimental research in the social sciences that deals with behavioral aspects. However, more importantly, note that we are not primarily interested in the prediction of our outcome variables, but in the (marginal) effect of AULP on the outcome variables. In other words, we are not claiming that AULP is the best predictor of contact and response rates, but that the marginal effect of AULP is statistically significant on both these rates.

[6]Specifically, the analysis that follows, with the results in Table 3, thus is based on 63–65 countries (of the 124 countries in Table 2) and corresponding numbers of unique country scores, depending on the availability of the culture scores for the countries in our sample and the overlap in availability when both scores are included in the same regression.

[7]The fixed effects specification results are available on request.

Briefly, estimates of Models 6 and 7 (described in Subsection 3.8) reveal that: AULP does not lead to unwanted side effects in terms of non-response bias (Appendix D.1); our results for time variables, filled items ratio, recontact (Appendices D.2 and D.3), as well as answer patterns, break-offs, and straightlining (Appendix D.4) do not reveal any evidence of adverse effects on data quality induced by the AULP treatment. Additional analysis using cooperation rates (Appendix D.1) also show no effect of the AULP treatment on cooperation rates. All the variables used for data quality are described and summarized in Appenices D.4 and D.6.

5 Discussion and Conclusion

This study explores whether announcing undefined lottery prizes in the subject line of survey invitation emails—abbreviated as AULP—increases contact and response rates, by randomizing the email addresses of a global sample of key decision-makers in MFIs into an AULP treatment group and a control group. Our analysis yields four main results.

First, we demonstrate that on average, AULP significantly increases contact and response rates. The effects are not only statistically significant but also are quantitatively meaningful: The contact rate increases by 9 percentage points, compared with an average contact rate of 37 percentage points in the control group, which is an increase of 25%. The response rate increases by 2 percentage points, and compared with an average response rate of 11 percentage points in the control group, this value amounts to an increase of 18%. We interpret these results as the AULP treatment increasing the salience of the lottery incentive, in line with the leverage-salience theory of survey participation (Groves et al., 2000).[8]

Second, we explore heterogeneity in the impact of the AULP treatment across different subgroups, such as OIC member states versus non-member states and, translated versus non-translated subgroups within the OIC and non-OIC states. These results reveal that AULP has a weaker treatment effect on contact rates in the OIC subgroup than in the non-OIC subgroup, which we interpret as a more negative leverage being assigned to lottery-style prize allocation mechanisms in OIC member states, possibly due to negative attitudes towards lottery prizes stemming from Islamic religious beliefs.[9] This result appears unremarked on by prior survey methodology literature, representing a novel contribution of our paper. Notably, it also is robust to using other control variables for culture, multi-level modeling, and country fixed effects. To the best of our knowledge, we provide the first analysis of the differential impact of making lottery incentives more salient (for a survey invitation email specifically and an online survey format more generally) in this context.

Third, in the non-OIC subgroup, the treatment effect on contact rates is stronger for the translated subgroup; similarly, the weaker AULP treatment effect in OIC countries is driven by the translated subgroup in these countries. These results can be interpreted as translations making AULP more salient, accentuating the more negative leverage attached to the underlying lottery in OIC countries.

Fourth, we do not find evidence that the AULP treatment increases non-response bias. People who open the invitation email after receiving the AULP treatment are not significantly different in terms of personal, MFI, or regional characteristics from those who open the invitation email without receiving AULP. Thus, we do not expect different responses to the incentive across groups. In fact, further analysis (reported in Appendix D.1), fails to reject the hypothesis that Cooperation Rate 1 differs significantly between treatment and control groups. In other words, the AULP effect does not arise because people in the treatment group who opened the survey invitation email reacted differently to the specific prizes than those in the control group. This tentatively suggests (at least for our context) that there is no reason for survey practitioners who want to use the AULP method to choose prizes that are qualitatively or quantitatively similar to ours. We leave it to future research to investigate this claim using different incentives in other settings.

Fifth, in evaluating the possible side effects on various other metrics of data quality (e.g., time variables, filled items ratio, recontact, answer patterns, break-offs, and straightlining), we find no evidence of adverse effects on data quality induced by the AULP treatment.

In general, because manipulating the subject line is costless, there are no incremental costs associated with implementing this strategy, and AULP should be implemented whenever the data show a positive effect on contact and response rates, without adverse effects on data quality. Thus, our results show that AULP should be implemented in non-

[8] These findings represent those of our baseline model without the inclusion of interaction terms for heterogeneity analyses. Our substantive conclusions about the effect of our treatment on contact rates is insensitive to the inclusion of these interaction terms. Some results regarding the treatment effect on response rates, however, lose statistical significance with the inclusion of these interaction terms, most likely due to the relatively low cooperation rate in our sample.

[9] The reader may be concerned that our conclusion about the OIC effect is influenced by the fact that the average response rate in OIC countries is higher. However, our results show that the lower AULP effect in the OIC subgroup (compared to non-OIC) is driven by the translated subgroup. This result corroborates our culture-based interpretation of the result since it is not clear why the OIC countries having a higher unconditional response rate would make the AULP effect lower in the translated subgroup of OIC countries only. Note also that the OIC dummy variable (by itself) controls for the response rate being higher on average in OIC countries; and the interaction term tells us that the AULP effect is lower in OIC countries after controlling for the response rate being higher on average in OIC countries.

Table 3
Sensitivity analyses: Further results on the role of culture

	Model 4 Model 3 with multi-level modeling		Model 5a Individualism		Model 5b Trust		Model 5c Individualism and trust	
Variable	(1) Contact[g]	(2) Response[h]	(3) Contact[g]	(4) Response[h]	(5) Contact[g]	(6) Response[h]	(7) Contact[g]	(8) Response[h]
AULP[a]	0.063**	0.029*	0.005	0.025	0.145***	0.029	0.128*	0.017
	(0.028)	(0.018)	(0.057)	(0.034)	(0.049)	(0.030)	(0.071)	(0.043)
OIC[c]	−0.022	−0.023	−0.040	−0.048	−0.036	−0.028	−0.039	−0.040
	(0.054)	(0.035)	(0.073)	(0.048)	(0.072)	(0.044)	(0.074)	(0.044)
AULP×OIC	0.038	−0.024	0.058	−0.004	0.055	−0.009	0.060	−0.005
	(0.059)	(0.041)	(0.076)	(0.050)	(0.074)	(0.047)	(0.076)	(0.048)
Transl[b]	−0.008	0.007	−0.047	−0.003	−0.048	−0.018	−0.051	−0.025
	(0.036)	(0.022)	(0.046)	(0.028)	(0.046)	(0.027)	(0.047)	(0.026)
AULP×Transl	0.093**	−0.013	0.109**	−0.024	0.096**	−0.024	0.094**	−0.020
	(0.036)	(0.023)	(0.044)	(0.028)	(0.044)	(0.028)	(0.044)	(0.028)
Transl×OIC	0.166**	0.060	0.184**	0.071	0.188**	0.083	0.189**	0.087*
	(0.065)	(0.041)	(0.085)	(0.053)	(0.084)	(0.050)	(0.086)	(0.049)
AULP×Transl×OIC	−0.139**	0.025	−0.164*	0.002	−0.155*	0.007	−0.152*	−0.001
	(0.069)	(0.046)	(0.085)	(0.055)	(0.085)	(0.054)	(0.085)	(0.054)
Individualism[e]	-	-	−0.001	−0.002	-	-	−0.0004	−0.002
	-	-	(0.002)	(0.001)	-	-	(0.002)	(0.001)
AULP×Individualism	-	-	0.002	0.001	-	-	0.001	0.001
	-	-	(0.002)	(0.001)	-	-	(0.002)	(0.001)
Trust[f]	-	-	-	-	0.001	−0.001	0.001	−0.001
	-	-	-	-	(0.001)	(0.001)	(0.001)	(0.001)
AULP×Trust	-	-	-	-	−0.003***	0.0002	−0.003***	0.0002
	-	-	-	-	(0.001)	(0.001)	(0.001)	(0.001)
Controls[d]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5128	5128	3524	3524	3550	3550	3473	3473
Number of groups	124	124	65	65	65	65	63	63

Sensitivity analyses with measures of cultural characteristics from Beugelsdijk and Welzel (2018). Marginal effects computed at the sample means from the multi-level logistic regressions, with standard errors in parentheses. See Table B1 in the Appendix for data sources.
[a] Dummy variable for potential respondent is (randomly) provided with an announcement of undefined lottery prizes in the subject line of the survey invitation emails. [b] Dummy variable for potential respondent received a survey translated into an official language of their country and 0 otherwise (see Appendix A.3 for criteria).
[c] Dummy variable for potential respondent is a staff member at an MFI is operating in an OIC member state.
[d] Dummy variables for target market of an MFI is broad/high end and for MFI is older than 8 years. [e] Score (0 to 100), reflecting the degree to which people in a country consider themselves autonomous personalities (higher score) or members of close communities (lower score). [f] Score (0-100), reflecting the degree to which people in a country feel comfortable in unstructured situations (higher score) or stressed and anxious in such situations (lower score). [g] Dummy variable for potential respondent opened the survey invitation email [h] Dummy variable for potential respondent fully completed the survey.
* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

OIC countries, though it is not recommended as *standard* practice for OIC countries due to the negative leverage effect discussed above. In fact, the role of lotteries in generating our results for OIC countries deserves further research consideration; for example, might the announcement of undefined *non-lottery* rewards be more effective in OIC countries?

Overall, we contribute to the literature studying the effects of manipulation of the subject line of survey invitation emails in the context of lottery prizes, where the papers closest to ours are Janke (2014) and Zhang et al. (2016). Similar to Janke (2014) and Zhang et al. (2016), our results show that any mention of lottery prizes in the subject line improves response rates without any negative effects on data quality. Janke (2014) reports an increase in overall response rates between his initial and final two survey rounds of 5 percent and 1 percent, respectively, subject to the caveats we noted in Subsection 2.2. These caveats include using the notion of "valid surveys received" instead of fully completed surveys; possible cohort effects and omitted variable bias; a problematic baseline; and the non-comparability of different samples, all of which make it difficult to compare our results with Janke (2014). We find an increase in the response rate of about 2 percentage points, and much larger increases in contact rates—which Janke (2014) does not study—of about 10 percentage points, even though we use the most conservative measures of both contact and response rates according to AAPOR 2016 standard definitions. Zhang et al. (2016), comparing their lottery-centered and survey-centered treatments, find a 5-percentage point higher response rate in the former, especially for relatively low-income participants, along with small, mostly statistically insignificant, adverse effects on data quality. However, when interpreting these numbers, it is important to keep in mind that Janke (2014) uses a sample of mostly university students, and Zhang et al. (2016) use a sample of employees of the institution sponsoring the survey. Notably, earlier studies have shown that student and employee populations tend to exhibit higher survey participation rates than the general population (Heberlein & Baumgartner, 1978; LaRose & Tsai, 2014; Peterson, 2001; Petrovcic, Petric, & Manfreda, 2016; Shih & Fan, 2008).

In contrast, we use a sample of real-world MFI practitioners, which also is pertinent when considering the generalizability of the results. Compared with top management in other financial institutions such as banks, the top management in MFIs tends to be relatively more heterogeneous, as reflected in the diversity of the institutions they represent, as well as their varied backgrounds. Mersland, Randoy, and Strom (2011) note that MFIs display a wide range of organizational and legal forms, such as non-profits (NGOs), cooperatives/credit unions, non-banking financial institutions, and banks. They point out that the first MFIs were organized as trusts/foundations, or non-profits, or were controlled by international non-profits, and the entry of commercial banks took place at a later stage. According to Mersland et al. (2011), the earliest MFIs were founded mostly by "social entrepreneurs, coming from a philanthropic development culture", and this characterization still appears true, at least to some degree (Beisland et al., 2019; Randoy, Strom, & Mersland, 2015). Partly due to this, more CEOs of MFIs have dual roles, in the management and the board of directors, and consequently exercise more control over the firm (Galema, Lensink, & Mersland, 2012). Randoy et al. (2015) refer to them as "motivated agents," in the sense of Besley and Ghatak (2005), such that they might be less motivated by high-powered incentives or less in need of costly corporate governance mechanisms to monitor their performance, in contrast to people in comparable positions in other financial firms (Deutsch, Keil, & Laamanen, 2011). In fact, despite some shifts in the MFI industry toward more shareholder-based ownership, most MFIs remain driven by their social aims (Armendariz & Morduch, 2010; Servin, Lensink, & van den Berg, 2012); their managers may be more akin to managers of hospitals or schools, as opposed to commercial banks (Galema et al., 2012). With regard to their business education, Pascal, Mersland, and Mori (2017) also note that MFI CEOs are more likely to come from varied educational backgrounds, not just business education. Finally, women are more prevalent among top MFI management (Périlleux & Szafarz, 2015; Strom et al., 2014). For example, Strom et al. (2014) report, for a sample of 329 MFIs in 73 countries from 1998 to 2008, a much larger proportion of top management who are women (27%, 23%, and 29% of CEOs, board chairs, board members) than in a typical U.S. firm (e.g., Adams and Ferreira (2009), find only 9% of directors being female). Considering this evidence of the diversity of the top management in MFIs, our results should apply widely. However, it remains important to implement our AULP treatment for different samples, to be able to more confidently generalize our findings to other contexts.

Regarding the magnitude of our effects, some further points are noteworthy. As a subject line manipulation, AULP directly affects people's decision to open the survey invitation email (captured by the contact rate). However, conditional on opening the survey invitation email, for a given individual, survey *completion* depends on a multitude of factors, including importantly: the design and content of the main body of the survey invitation email, the nature of the prizes, as well as the design and nature of the survey itself, with consequent response rates being smaller than or equal to contact rates by definition. Given the specifics of the research question that we addressed with the questionnaire (which is not part of this paper), the inclusion of several questions related to financial and operational details of the respondents' organization into our questionnaire was required. This, combined with the fact that the only contact information we could obtain was often that of a senior executive (who might not necessarily have been the person most suited to provide the data we wanted), suggests that the relatively large difference between contact and response rates is not surprising in the context of our study. In general, AULP can be implemented with surveys of a different nature as well as different samples, including in contexts with much less of an expected discrepancy between the contact and response rates. For these reasons we think that it is important to present results for both contact and response to be able to support comparisons with other results from the literature.

Our results suggest several additional avenues for future research. In general, the substantial global coverage in our sample (124 countries) came at the cost of limited availability of data about the demographic and socio-economic characteristics of the email recipients. We thus cannot explore heterogeneity in our AULP treatment effect along important dimensions such as age, education, and income level. We believe it would be instructive to explore additional variables such as income levels, which might be important predictors of whether treatments that build on lottery incentives perform well (Zhang et al., 2016). Our results pertaining to the group of OIC countries are novel and should be explored using more granular data for individual countries and respondents. The results of our non-response analysis show that for our sample, for a small number of characteristics (2 out of 19), the AULP treatment actually alleviates existing non-response bias. Further explorations of the effect of AULP on non-response bias, with a richer set of demographic characteristics and different samples, would be of interest.

In addition, while our AULP treatment was randomized, translations were not. Including additional randomized treatments in which survey invitation emails and questionnaires are sent to two random subsamples, with and without offering translations, would allow for a more rigorous interpretation of translation effects and also permit meaningful cost-effectiveness calculations for translations.

We also find that email reminders increase response rates substantially. However, the overuse of email reminders in the context of AULP, in terms of sending more reminders in the same time span may actually reduce response rates; the same goes for sending more reminders over a longer period of time, which would lead to a delayed disbursement of prizes, and impatient prospective respondents thus discounting prizes more heavily. Such trade-offs should be explored further. These and other investigations are left for further work.

References

Adams, R., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, *94*(2), 291–309.

Ahmad, S., Lensink, R., & Mueller, A. (2020). The double bottom line of microfinance: A global comparison between conventional and Islamic microfinance. *World Development*, Forthcoming.

Armendariz, B., & Morduch, J. (2010). *The economics of microfinance*. London: The MIT Press.

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, *61*, 1139–1160.

Beisland, L. A., Ndaki, D. P., & Mersland, R. (2019). Influence of ownership type and CEO power on residual loss: Evidence from the global microfinance industry. *Nonprofit and Voluntary Sector Quarterly*.

Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review*, *95*, 616–636.

Beugelsdijk, S., & Welzel, C. (2018). Dimensions and dynamics of national culture: Synthesizing Hofstede with Inglehart. *Journal of Cross-Cultural Psychology*, *49*, 1469–1505.

Birnholtz, J. P., Horn, D. B., Finholt, T. A., & Bae, S. J. (2004). The effects of cash, electronic, and paper gift certificates as respondent incentives for a web-based survey of technologically sophisticated respondents. *Social Science Computer Review*, *22*, 355–362.

Bonke, J., & Fallesen, P. (2010). The impact of incentives and interview methods on response quantity and quality in diary- and booklet-based surveys. *Survey Research Methods*, *4*, 91–101.

Boulianne, S., Klofstad, C. A., & Basson, D. (2011). Sponsor prominence and responses patterns to an online survey. *International Journal of Public Opinion Research*, *23*.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, *57*, 62–79.

Cloud, I. M. (2016). *Email marketing metrics benchmark study*.

Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement*, *60*, 821–836.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, *64*, 464–494.

Couper, M. P. (2001). Web surveys: The questionnaire design challenge.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, 87–98.

Daikeler, J., Bosnjak, M., & Lozar Manfreda, K. (2019). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology*, 1–27.

de Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In *Survey nonresponse* (pp. 41–54). New York: Wiley.

Delhey, J., & Newton, K. (2005). Predicting cross-national levels of social trust: Global pattern or Nordic exceptionalism? *European sociological review*, *21*(4), 311–327.

Deutsch, Y., Keil, T., & Laamanen, T. (2011). A dual agency view of board compensation: The joint effects of outside director and CEO stock options on firm risk. *Strategic Management Journal*, *32*, 212–227.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. Wiley-Interscience, New York.

Dillman, D. A., & Bowker, D. K. (2001). The web questionnaire challenge to survey methodologists. In *Dimen-*

sions of Internet science. Lengerich, Germany: Pabst Science Publishers.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd). New York: NY: John Wiley & Sons.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th). New York, United States: John Wiley & Sons Inc.

ESOMAR. (2018). *Global market research report 2018: An ESOMAR industry report*. ESOMAR.

Fowler, F. J. (2012). *Survey research methods*. Thousand Oaks, CA: Sage Publications.

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly*, *52*, 467–491.

Galema, R., Lensink, R., & Mersland, R. (2012). Governance and microfinance institutions. In J. R. Barth, C. Lin, & C. Wihlborg (Eds.), *Research handbook on international banking and governance*. UK: Edward Elgar.

Goeritz, A. (2006). Incentives in web studies: Methodological issues and a review. *International Journal of Internet Science*, *1*(1), 58–70.

Goeritz, A., & Luthe, S. (2013a). Effects of lotteries on response behavior in online panels. *Field Methods*, *25*(3), 219–237.

Goeritz, A., & Luthe, S. (2013b). Lotteries and study results in market research online panels. *International Journal of Market Research*, *55*(3), 611–626.

Goyder, J. (1982). Further evidence on factors affecting response rates to mailed questionnaires. *American Sociological Review*, *47*, 550–553.

Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the effects of removing "too fast" responses and respondents from web surveys. *Public Opinion Quarterly*, *79*(2), 471–503.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, *56*, 475–495.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias; a meta analysis. *Public Opinion Quarterly*, *72*, 167–189.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an illustration. *Public Opinion Quarterly*, *64*, 299–308.

Hansen, K. M. (2006). The effects of incentives, interview length, and interviewer characteristics on response rates in a CATI-study. *International Journal of Public Opinion Research*, *19*, 112–121.

Harkness, J. A., van de Vijver, F. J., & Mohler, P. P. (2002). *Cross-cultural survey methods*. Wiley Series in Survey Methodology.

Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, … T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 117–140). John Wiley & Sons.

Harzing, A.-W. (2000). Cross–national mail surveys: Why do response rates differ between countries? *Industrial Marketing Management*, *29*, 243–254.

Hassanat, A. B., & Al tarawneh, G. A. (2015). Gambling-free lottery, a new Islamic lending ticket. *Journal of Islamic Accounting and Business Research*, *6*, 42–54.

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, *43*, 447–462.

Heerwegh, D. (2006). An investigation of the effect of lotteries on web survey response rates. *Field Methods*, *18*, 205–220.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind.* (3rd ed.).

Hoonakker, P., & Carayon, P. (2013). Questionnaire survey nonresponse: A comparison of postal mail and internet surveys. *International Journal of Human-Computer Interaction*, *25*, 348–373.

Jaeckle, A., Burton, J., Couper, M. P., & Lessof, C. (2019). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Coverage and participation rates and biases. *Survey Research Methods*, *13*(1), 23–44.

Janke, R. (2014). Effects of mentioning the incentive prize in the email subject line on survey response. *Evidence Based Library and Information Practice*, *9*.

Joinson, A. N., & Reips, U.-D. (2007). Personalized salutation, power of sender and response rates to web-based surveys. *Computers in Human Behavior*, *23*, 1372–1383.

Kaplowitz, M. D., Lupi, F., & Arreola, O. (2012). Local markets for payments for environmental services: Can small rural communities self-finance watershed protection? *Water Resources Management*, *26*, 3689–3704.

Kluckhohn, F. R., & Strodtbeck, F. L. (1961). Variations in value orientations.

Kypri, K., & Gallagher, S. J. (2003). Incentives to increase participation in an internet survey of alcohol use: A controlled experiment. *Alcohol and Alcoholism*, *38*, 437–441.

Laguilles, J. S., Williams, E. A., & Saunders, D. B. (2011). Can lottery incentives boost web survey response rates? Findings from four experiments. *Research in Higher Education*, *52*, 537–553.

LaRose, R., & Tsai, H.-y. S. (2014). Completion rates and non-response error in online surveys: Comparing sweepstakes and pre-paid cash incentives in studies of online behavior. *Computers in Human Behavior*, *34*, 110–119.

Lazar, J., & Preece, J. (1999). Designing and implementing web-based surveys. *Journal of Computer Information Systems*, *39*, 63–67.

Lozar Manfreda, K., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, *50*, 79–104.

Marcus, B., Bosnjak, M., Lindner, S., Pilischenko, S., & Schuetz, A. (2007). Compensating for low topic interest and long surveys: A field experiment on nonresponse in web surveys. *Social Science Computer Review*, *25*.

Martin, C. L. (1994). The impact of topic interest on mail survey response behavior. *Journal of the Market Research Society*, *36*, 327–337.

Mazareanu, E. (2019). Countries with the largest market research revenue worldwide 2018. In *Market research*. Statista.

Mersland, R., Randoy, T., & Strom, R. (2011). The impact of international influence on microbanks' performance: A global survey. *International Business Review*, *20*, 163–176.

Morea, S. (2014). *Market research in the US*. IBIS World Industry Report 54191.

Parsons, N. L., & Manierre, M. J. (2014). Investigating the relationship among prepaid token incentives, response rates, and nonresponse bias in a web survey. *Field Methods*, *26*, 191–204.

Pascal, D., Mersland, R., & Mori, N. (2017). The influence of the CEO's business education on the performance of hybrid organizations: The case of the global microfinance industry. *Small Business Economics*, *49*, 339–354.

Patrick, M. E., Singer, E., Boyd, C. J., Cranford, J. A., & McCabe, S. E. (2013). Incentives for college student participation in web-based substance use surveys. *Search Results*, 1710–1714.

Périlleux, A., & Szafarz, A. (2015). Women leaders and social performance: Evidence from financial cooperatives in senegal. *World Development*, *74*, 437–452.

Petchenik, J., & Watermolen, D. J. (2011). A cautionary note on using the internet to survey recent hunter education graduates. *Human Dimensions of Wildlife*, *16*, 216–218.

Peterson, P. W. (2001). The debate about online learning: Key issues for writing teachers. *Computers and Composition*, *18*, 359–370.

Petrovcic, A., Petric, G., & Manfreda, K. L. (2016). The effect of email invitation elements on response rate in a web survey within an online community. *Computers in Human Behavior*, *56*, 320–329.

Pforr, K. (2016). Incentives. In *Gesis survey guidelines*. Mannheim, Germany: GESIS—Leibniz Institute for the Social Sciences.

Porter, M. E. (2004). *Competitive advantage: Creating and sustaining superior performance*. Free Press.

Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey reponse rates. *Research in higher education*, *44*, 389–407.

Porter, S. R., & Whitcomb, M. E. (2005). E-mail subject lines and their effect on web survey viewing and response. *Social Science Computer Review*, *23*, 380–387.

Randoy, T., Strom, R. O., & Mersland, R. (2015). The impact of entrepreneur-ceos in microfinance institutions: A global survey. *Entrepreneurship Theory and Practice*, 1042–2587.

Rogelberg, S. G., & Stanton, J. M. (2007). Introduction: Understanding and dealing with organizational survey nonresponse. *Organizational Research Methods*, *10*, 195–209.

Sánchez-Fernández, J., Muñoz-Leiva, F., & Montoro-Rios, F. J. (2012). Improving retention rate and response quality in web-based surveys. *Computers in Human Behavior*, *28*(2), 507–514.

Sappleton, N., & Lourenco, F. (2016). Email subject lines and response rates to invitations to participate in a web survey and a face-to-face interview: The sound of silence. *International Journal of Social Research Methodology*, *19*, 611–622.

Servin, R., Lensink, R., & van den Berg, M. (2012). Ownership and technical efficiency of microfinance institutions: Empirical evidence from Latin America. *Journal of Banking and Finance*, *36*, 2136–2144.

Sheehan, K. B., & McMillan, S. J. (1999). Response variation in email surveys: An exploration. *Journal of Advertising Research*, *39*, 45–54.

Shih, T.-H., & Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. *Field Methods*, *20*, 249–271.

Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*, 112–141.

Strom, R. O., D'Espallier, B., & Mersland, R. (2014). Female leadership, performance, and governance in microfinance institutions. *Journal of Banking and Finance*, *42*, 60–75.

Toninelli, D., & Revilla, M. (2016). Smartphones vs PCs: Does the device affect the web survey experience and the measurement error for sensitive topics? A replication of the Mavletova & Couper's 2013 experiment. *Survey Research Methods*, *10*, 153–169.

Triandis, H., & Gelfand, M. (2012). A theory of individualism and collectivism. In L. PAM, A. Kruglanski, & E. Higgins (Eds.), *Handbook of theories of social psychology 2* (pp. 498–520). London: Sage.

Trouteaud, A. R. (2004). How you ask counts: A test of internet-related components of response rates to a web-based survey. *Social Science Computer Review*, *22*(3), 385–392.

Tyupa, S. (2011). A theoretical framework for back-translation as a quality assessment tool. *New Voices in Translation Studies*, *7*, 35–46.

Van Horn, P. S., Green, K. E., & Martinussen, M. (2009). Survey response rates and survey administration in counseling and clinical psychology: A meta-analysis. *Educational and Psychological Measurement*, *69*, 389–403.

Van Mol, C. (2017). Improving web survey efficiency: The impact of an extra reminder and reminder content on web survey response. *International Journal of Social Research Methodology*, *20*, 317–327.

Walsh, J. P., Kiesler, S., Sproull, L. S., & Hesse, B. W. (1992). Self-selected and randomly selected respondents in a computer network survey. *Public Opinion Quarterly*, *56*, 21–244.

Walston, J. T., Lissitz, R. W., & Rudner, L. M. (2006). The influence of web-based questionnaire presentation variations on survey cooperation and perceptions of survey quality. *Journal of Official Statistics*, *22*, 271–291.

Warriner, K., Goyder, J., Gjertsen, H., Hohner, P., & McSpurren, K. (1996). Charities, no; lotteries, no; cash, yes: Main effects and interactions in a canadian incentives experiment. *Public Opinion Quarterly*, *60*, 542–562.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior: A meta-analysis. *Public Opinion Quarterly*, *55*, 613–639.

Young, M. (2007). *More to Islamic branding than meets the eye*. UK: Campaign.

Zak, P. J., & Knack, S. (2001). Trust and growth. *The economic journal*, *111*(470), 295–321.

Zhang, C., Lonn, S., & Teasley, S. D. (2016). Understanding the impact of lottery incentives on web survey participation and response quality—a leverage-salience theory perspective. *Field Methods*, *29*.

Appendix A
Study design information

A.1 List of states for which translations were provided

Note: Highlighted states are OIC member states.

Arabic and English (206 emails) *Bahrain*, *Egypt*, *Iraq*, Israel, *Jordan*, *Lebanon*, *Morocco*, *Palestine*, *Saudi Arabia*, *Sudan*, *Syria*[10], *Tunisia*, *Yemen*.

Bengali and English (198 emails) *Bangladesh*.

Chinese and English (32 emails) China.

French and English (719 emails) *Benin*, *Burkina Faso*, Burundi, *Cameroon*, Central African Republic, *Côte d'Ivoire*, Democratic Republic of the Congo, *Gabon*, *Guinea*, Haiti, Madagascar, *Mali*, *Niger*, Republic of the Congo, Rwanda, *Senegal*, *Togo*, Vanuatu.

Hindi and English (537 emails) India.

Indonesian and English (59 emails) *Indonesia*.

Russian and English (398 emails) *Azerbaijan*, Belarus, *Kazakhstan*, *Kyrgyzstan*, Russia.

Spanish and English (1,031 emails) Argentina, Bolivia, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, Venezuela.

Urdu and English (158 emails) *Pakistan*.

A.2 List of states for which no translation was provided

Note: Highlighted states are OIC member states.

English only (1,790 emails) *Afghanistan*, *Albania*, Angola, Armenia, Belize, Bhutan, Bosnia and Herzegovina, Brazil, Bulgaria, Cambodia, Croatia, East Timor, Ethiopia, Fiji, *Gambia*, Georgia, Ghana, Grenada, *Guinea-Bissau*, *Guyana*, Hungary, Jamaica, Kenya, Kosovo, Laos, Liberia, Macedonia, Malawi, *Malaysia*, Moldova, Mongolia, Montenegro, *Mozambique*, Myanmar, Namibia, Nepal, *Nigeria*, Papua New Guinea, Philippines, Poland, Portugal, Romania, Saint Lucia, Samoa, Serbia, *Sierra Leone*, Slovakia, Solomon Islands, South Africa, South Sudan, Sri Lanka, *Suriname*, Swaziland, *Tajikistan*, Tanzania, Thailand, Tonga, Trinidad and Tobago, *Turkey*, *Uganda*, Ukraine, United States, *Uzbekistan*, Vietnam, Zambia, Zimbabwe.

A.3 Translation criteria

First, countries were grouped into different language groups based on their most commonly spoken official language (besides English if English was an official language for a country). For example, India was classified into the Hindi language group because even though India has 24 official languages including English and Hindi, the latter is the most commonly spoken.

Second, we determined whether a *language group*, which as explained above could have had more than one *country* in it, fulfilled two simple criteria:

1. There were more than 70 MFIs of our dataset located in countries that were assigned to that language group.

2. The language group had over 50 million native speakers in all the countries we classified under that language group combined, i.e., it was a widely spoken language globally. This determination was made using Ethnologue data for 2015 and 2019, in combination with data from the World Factbook 2015.

Table A1 summarizes whether these two criteria are fulfilled, for each country in our sample.

Third, we provided a translation for each *language group* that met *both* of the above criteria simultaneously. Note, however, that we made four exceptions (for Chinese, Indonesian, Portuguese and Urdu) to the above rule:

1. We provided a translation into Urdu, Indonesian, and Chinese even though they did not satisfy the number of MFIs criterion. The reasons for this are as follows. Pakistan and Indonesia are two of the countries that the literature on Islamic MFIs focuses on the most, and we wanted to be able to speak to this literature when discussing our survey's content/analysis. In the case of Chinese, it was for cost-benefit reasons. The benefit came from its size and importance in the world economy (when discussing our survey's content/analysis), while the cost of translation was comparatively affordable because our university has an in-house translation resource which translated Chinese.

2. We did not provide a translation into Portuguese even though it qualified for translation, due to logistical and budgetary reasons, mainly because our university's in-house translation resource does not offer Portuguese translation.

Table A1 presents the translation decisions for each country in our sample.

A.4 Sample selection and randomization procedure

Step1 We obtained a spreadsheet from MIX Market with 5,649 rows of data containing contact details of individual staff members of MFIs (who were listed as designated contact persons of these MFIs). Note that at

[10] Syria joined the OIC in 1970. At the time that this paper was written, Syria's membership of the OIC remained (temporarily) suspended.

Table A1
Provision of translations

Language Group	Country	MFIs >70	Native speakers >50 Million	Translation Decision
Albanian	Albania, Kosovo	X	X	X
Amharic	Ethiopia	X	X	X
Arabic	Bahrain, Egypt, Iraq, Israel, Jordan, Lebanon, Morocco, Palestine, Saudi Arabia, Sudan, Syria, Tunisia, Yemen	✓	✓	✓
Armenian	Armenia	X	X	X
Bengali	Bangladesh	✓	✓	✓
Bulgarian	Bulgaria	X	X	X
Burmese	Myanmar (Burma)	X	X	X
Chinese	China	X	✓	✓
Croatian	Bosnia and Herzegovina, Croatia	X	X	X
Dutch	Suriname	X	X	X
Dzonkha	Bhutan	X	X	X
English (only)	Belize, Ghana, Grenada, Guyana, Jamaica, Liberia, Malawi, Namibia, Nigeria, Saint Lucia, Sierra Leone, Solomon Islands, South Sudan, The Gambia, Trinidad and Tobago, United, States, Zambia	✓	✓	English is de jure language
Fijian	-	X	X	X
Filipino	Philippines	✓	X	X
French	Benin, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Comoros, Côte d'Ivoire, Democratic Republic of the, Congo, Gabon, Guinea, Haiti, Madagascar, Mali, Niger, Republic of the Congo, Rwanda, Senegal, Togo, Vanuatu	✓	✓	✓
Georgian	Georgia	X	X	X
Hindi	India	✓	✓	✓
Hiri Motu	Papua New Guinea	X	X	X
Hungarian	Hungary	X	X	X
Indonesian	Indonesia	X	✓	✓
Khmer	Cambodia	X	X	X
Lao	Laos	X	X	X
Macedonian	Macedonia	X	X	X
Malay	Malaysia	X	X	X
Mongolian	Mongolia	X	X	X
Montenegrin	Montenegro	X	X	X
Nepali	Nepal	X	X	X
Pashto	Afghanistan	X	X	X
Polish	Poland	X	X	X
Portuguese	Angola, Brazil, East Timor, Guinea-Bissau, Mozambique, Portugal	✓	✓	X
Romanian	Moldova, Romania	X	X	X
Russian	Azerbaijan, Belarus, Kazakhstan, Kyrgyzstan, Russia	✓	✓	✓
Samoan	Samoa	X	X	X

Continues on next page

Continued from last page

Language Group	Country	MFIs >70	Native speakers >50 Million	Translation Decision
Serbian	Serbia	X	X	X
Shona	Zimbabwe	X	X	X
Sinhala	Sri Lanka	X	X	X
Slovak	Slovakia	X	X	X
Spanish	Argentina, Bolivia, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, Venezuela	✓	✓	✓
Swahili	Kenya, Tanzania, Uganda	✓	X	X
Swati	Swaziland	X	X	X
Tajiki	Tajikistan	X	✓	X
Thai	Thailand	X	X	X
Tongan	Tonga	X	X	X
Turkish	Turkey	X	✓	X
Ukrainian	Ukraine	X	X	X
Urdu	Pakistan	X	✓	✓
Uzbek	Uzbekistan	X	X	X
Vietnamese	Vietnam	X	✓	X
Zulu	South Africa	X	X	X

this point we refer to rows of data for reasons that will become clear below.

Step 2 We dropped 351 rows of data that had incomplete or missing email addresses, leaving us with 5,298 rows of data with usable email addresses.

Step 3 We dropped 170 rows, which had duplicate email addresses. This could happen when, for example, the same MFI had two rows of data with two different contact persons but the same email address. This step left us with 5,128 rows of data with 5,128 unique and usable email addresses. Note that it is possible for an MFI to have multiple rows of data because MIX Market provided multiple contact persons, each with a unique email address, for the same MFI. We kept all such entries so that our 5,128 unique email addresses correspond to 2,641 unique MFIs.

Step 4 We then classified these 5,128 unique email addresses into 10 language groups:

English (1,790 email addresses) and 9 non-English language groups (e.g., Arabic, French, etc.; a total of 3,338 email addresses) according to the criteria listed in Subsection A.3. Note that our classification ensured that a country would only be placed into one of the language groups.

The 1,790 email addresses classified into the English group received the email only in English, and *all* the other 3,338 email addresses classified into the 9 non-English language group received the email in English as well as the translation into the group language.

Step 5 *Within each of the 10 language groups,* we randomized half of the email addresses into our treatment group and half into our control group. Thus, half of the 1,790 email addresses in the English group received the treatment. The same was done for each of the 9 other language groups, so that out of the total of 3,338 email addresses classified into the non-English group, half (1,699) were in the treatment group.

A.5 Details of randomization

The actual randomization was implemented as follows: we copied all the email addresses from a particular language group into a separate MS Excel sheet (i.e., a total of 10 separate sheets). We then used the random number generator function of MS Excel to generate a random number for each email address. These numbers and their corresponding email addresses were then sorted from high to low within each of the 10 sheets, i.e., for each of the 10 language groups. The top 50% of the random numbers (and their corresponding email addresses) were assigned to the treatment group, the rest to the control group.

Figure A1 visualizes the steps outlined above

A.6 Invitation email

Two different subject lines were used, as follows. The full exhibit of the invitation email is shown in Figure A2.

Subject line for treatment group "5-minute Survey on MFIs—Exciting Prizes."

Subject line for control group "5-minute Survey on MFIs."

A.7 Costs of incentives: expected costs, consumption of prizes and actual costs

This section provides a detailed breakdown of the costs of the lottery incentives that were employed in this study, including expected and actual costs, as well information on the consumption of prizes.

1. The lottery drawing actually took place and in general, all the information provided to respondents in the context of this study was accurate and truthful. We did not use deception in any way.

2. All respondents, irrespective of whether or not they received AULP, were eligible to win the following prizes. There was 1 first prize (the right to attend an international Summer School on microfinance at a top 100 research university, including all travel and lodging expenditures), which was consumed; there was 1 second prize (identical to the first prize, excluding travel expenditures), which was not consumed; there were 6 identical prizes in the third prize category (free access to instruction material used for the Summer School), which were consumed; and there were 10 identical prizes in the fourth prize category (donations of 50 Euros each to the MFIs where the winners in this prize-category were employed), of which only 3 were consumed. For the fourth prize it was necessary for the winners to send us the bank account information of their organization, which might have been partly responsible for the sharp differences in prize consumption between the third and fourth prize categories.

3. Anticipated versus actual costs for the different prize categories were as follows. The actual cost of the single first prize was € 2,612.17 which included items such as the air ticket (€ 1,467.17), the Summer School fee (€ 500), hotel (€ 250), meals (€ 210), and a number of miscellaneous items that added up to less than € 200. It was difficult ex-ante to anticipate the cost of the first prize, since the largest proportion consisted of an international flight round ticket, which would depend on who won the prize (e.g., the eventual winner was from Malawi, which did not have direct flights to the venue of the promised Summer School; this would have been different if the winner was from, say, South Africa or Dubai). However, it is worth noting that the actual flight price (€ 1,467.17) is fairly close to the upper limit of the flight price we anticipated because the route in question is one of the most expensive. Hence in this specific case, we use the actual flight price in the place of a maximum anticipated flight price.

The anticipated cost of the single second prize was € 500 and the actual cost was € 0 (because it was not consumed); the anticipated and actual costs of the third prize category were both € 0; and the anticipated cost of the entire fourth prize category was € 500 (€ 50 times 10, see point b) plus expected transfer fees, and the actual cost was € 150 (since only 3 out of 10 were consumed) plus actual transfer fees.

The total anticipated cost of incentives amounted to € 3,612.17, while the total actual cost of incentives amounted to 2,762.17. As a percentage of total anticipated and total actual costs these are 78% and 73%, respectively. Recall that everyone who got sent an invitation email was eligible for participating in the lottery; thus, these costs are not associated with the treatment we study in this paper, AULP, which is costless.

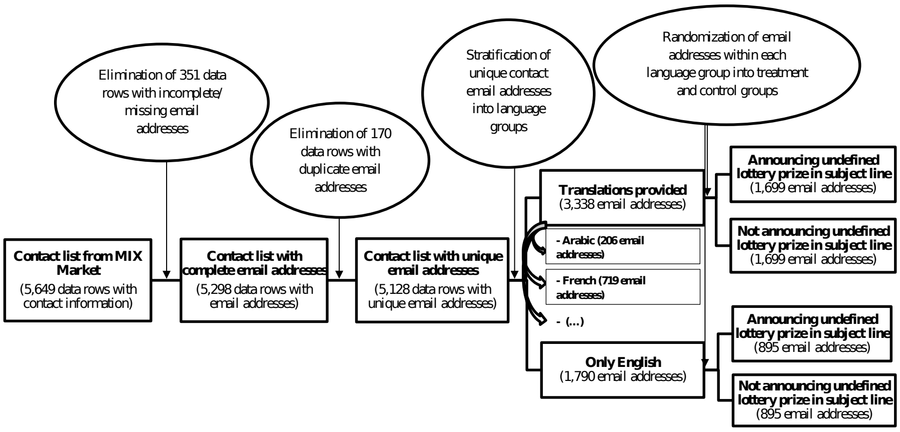

Figure A1. Flowchart for randomization procedure[1]

Table A2
Balancing test

	Overall	Treatment Group		Control Group		AULP Treatment-Control	
	Mean	Mean	Std. Dev.	Mean	Std. Dev.	Diff. T-C	p-value
Individuals' characteristics							
Male	0.713	0.722	0.448	0.704	0.456	0.017	0.174
Top management	0.538	0.536	0.499	0.499	−0.004	0.779	
MFIs' characteristics							-
Mature	0.569	0.584	0.493	0.555	0.497	0.03	0.032
Offices	45.693	42.141	184.465	49.245	22.196	−7.104	0.737
Sustainable	0.543	0.551	0.497	0.535	0.499	0.016	0.262
Regulated	0.610	0.606	0.489	0.614	0.487	−0.008	0.567
Profit margin	0.023	−0.014	0.861	0.059	4.879	−0.073	0.189
Female percentage	0.403	0.401	0.352	0.405	0.355	−0.004	0.774
Outreach	0.218	0.219	0.414	0.217	0.412	0.002	0.892
Scale	0.344	0.347	0.476	0.341	0.474	0.005	0.681
Target market	0.404	0.416	0.493	0.392	0.488	0.024	0.078
For profit	0.433	0.433	0.496	0.432	0.495	0.002	0.910
Regional characteristics							
English Official	0.342	0.344	0.475	0.341	0.474	0.003	0.837
Africa	0.268	0.262	0.440	0.273	0.446	−0.011	0.377
Asia Pacific	0.095	0.097	0.296	0.094	0.291	0.004	0.669
Europe Asia	0.156	0.161	0.368	0.151	0.358	0.011	0.299
America Caribbean	0.234	0.236	0.425	0.233	0.423	0.003	0.792
Middle East	0.040	0.040	0.196	0.040	0.196	0	1
South Asia	0.206	0.203	0.402	0.209	0.407	−0.006	0.581

Notes. This table presents "balancing tests", i.e., for each of the above variables, a statistical test for the difference in their means between the treatment and control group (along with their overall means; means for the treatment and control groups; and standard deviations). The variables used are individual characteristics of MFI practitioners, and associated MFIs and regions, for the full sample ($N = 5,128$). Treatment and control groups each have $N = 2,564$. For Offices, Profit margin and Female percentage, the reported p-values are for the difference between treatment and control group with the Wilcoxon rank-sum test. For all other characteristics, the reported p-value is for the difference between treatment and control group with the χ^2 test. Variable definitions are listed in Appendix B1. The unit of observation is an MFI practitioner associated with a unique MFI and location. Data source: Mix Market database and the authors' survey.

Dear Sir / Madam [Name]:

You are kindly invited to participate in a 5-minute research survey on microfinance institutions conducted by University XXXX in country XXX.

As a token of our appreciation for your participation in this study you can choose to enter into a lottery to win one of many exciting prizes - including full sponsorship of attendance at a microfinance themed Summer School at University XXXX, in year XXX year XXX (see l://SummerSchoolInformationLink for this year's Summer School). For a detailed description of all the prizes, please see below. l://DescriptionOfAllPrizes

The researchers conducting this study are principal investigator xxx, co-investigator 1, and co-investigator 2 from the University XXXX. If you have any questions about this study, you may contact principal investigator xxx at email address xxx or at contact number xxx. Alternatively, you may contact co-investigator 1 xxx at email address xxx or reach co-investigator 2 xxx at email address xxx. Your participation is highly valuable to us and will be much appreciated.

Taking part in this study is completely voluntary. If you decide to take part in the study, you may skip any question that you do not wish to answer. If you decide to skip some of the questions, it will not affect your chances of winning any of the prizes, nor will it affect your current or future relationship with the University XXXX. If you decide to take part in the study, you are also free to withdraw at any time.

All your answers will be kept confidential. In any report we make public, we will not include any information that will make it possible to identify you or your organization. The records of this study will be kept private, in a locked file to which only the researchers conducting this study will have access.
By clicking on the following link, I consent to participate in this research study.

Take the survey

Or copy and paste the URL below into your internet browser:

l://SurveyURL

Yours very sincerely,

Principal investigator xxx
Co-investigator 1 xxx
Co-investigator 2 xxx
University XXXX, City XXXX, Country XXXX
Date: XXXX

Follow the link to opt out of future emails:

http://rug.eu.qualtrics.com/CP/Register.php?OptOut=true&RID=MLRP_cYMecK2zIv90cFn&LID=UR_cv7SWGWn5Fhh6FT&BT=cnVn&_=1

Figure A2. Invitation Email

Appendix B
Information about key variables

B.1 Definitions of response, cooperation and contact rate

We calculate response rates according to AAPOR's 2016 standard definition of *Response Rate 1*. We calculate cooperation rate according to AAPOR's 2016 standard definition of *Cooperation Rate 1*. We calculate contact rate according to AAPOR's standard definition of *Contact Rate 1*. We use the figure B1 to simplify exposition. Note that $n(x)$ refers to the number of observations in a category x in the figure.

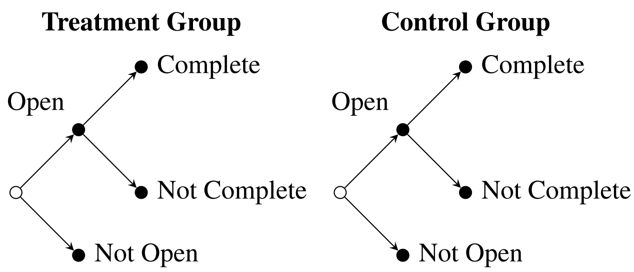

Figure B1. Graphical aid for understanding contact and response rate calculations

For both the treatment and control groups the definitions are as follows.

$$\text{Response rate} = \frac{n(\text{complete})}{n(\text{open}) + n(\text{not open})} = \frac{n(\text{complete})}{n(\text{sent})}$$

$$\text{Cooperation rate} = \frac{n(\text{complete})}{n(\text{open})}$$

$$\text{Contact rate1} = \frac{n(\text{open})}{n(\text{sent})}$$

with:

$$\begin{aligned} n(\text{complete}) &= 1.1 \\ n(\text{sent}) &= 1.1 + 1.2 + 2.111 + 2.1121 + 2.1122 + 2.12 \\ &\quad + 3.19 + 3.30 \\ n(\text{open}) &= 1.1 + 1.2 + 2.111 + 2.1121 + 2.1122 + 2.12. \end{aligned}$$

whereby the numbers in the right hand side refer to the AAPOR's final dispostion codes. two numbers in the last term, in accordance with AAPOR's 2016 standard definitions. Note that n(not open) is then naturally comprising all residual codes in n(sent), which implies that n(not open) = 3.19 + 3.30 in the context of our study.

B.2 Cultural factors

This section describes and motivates in more detail the two cultural indices, *Individualism* and*Trust*, used in our analysis. It also provides the individual country scores for these two indices.

Since our study was conducted globally, cultural factors are important to control for mainly because they might affect the patterns/rates of contact/response. We thus control for two cultural dimensions capturing beliefs about social structure *(Individualism)* and the nature of human behavior*(Trust)*, as explained by Kluckhohn and Strodtbeck (1961), in order to: check robustness of our results regarding OIC, and to make our results across countries comparable.

We briefly explain the rationale for using the specific measures of *Individualism* and *Trust*.

In general, *Individualism* explains the link between an individual in a society and collectivity (Hofstede, 2001), especially in terms of how autonomous or rooted within groups individuals are (Triandis & Gelfand, 2012). In our experiment, we employ the AULP treatment. Though gambling is in principle taboo in many religions, people's reaction to AULP in different countries will partly depend on the degree to which prevailing societal norms reflect religious norms, and how much importance they attach to religion. This aspect is captured by *Individualism*.

In general, social *Trust* explains the extent to which the individuals in a society feel comfortable facing uncertain circumstances and the degree of trust when interacting with others (Delhey & Newton, 2005; Zak & Knack, 2001). Individuals in societies with strong attitudes of avoidance towards uncertain situations may prefer tasks involving no or minimal risks. In our experimental context, this may be an important cultural aspect because societies with higher social trust may positively affect the potential respondent's attitude towards survey and lottery participation.

Country scores for cultural factors: We use wave-averaged country scores for *Individualism* and *Trust* from Beugelsdijk and Welzel (2018), who replicate Hofstede, Hofstede, and Minkov (2010) cultural dimensions scores for the largest possible sample of countries by using integrated survey data from the "World Values Survey" and the "European Values Survey". Table B2 shows the scores for the countries in our dataset.

Table B1
Variable definitions and sources

Variable	Abbreviation	Definition
The source of the following variables is the authors' survey		
Announcing Undefined Lottery Prizes	AULP	A dummy variable that takes the value of 1 if a potential respondent is (randomly) provided with the announcement of undefined lottery prizes in the subject line of the survey invitation emails; 0 otherwise.
Translated	Transl	A dummy variable that takes the value of 1 if a potential respondent received a survey translated into an official language of their country; 0 otherwise (see Appendix A.3 for criteria).
Contact		A dummy variable that takes the value of 1 if a potential respondent opened the survey invitation email; 0 otherwise.
Response		A dummy variable that takes the value of 1 if a potential respondent fully completed the survey; 0 otherwise.
The source of the following variables is MIX Market		
Male	Male	A dummy variable that takes the value of 1 if the potential respondent is male; 0 otherwise.
Potential respondent in a top management team	Top management	A dummy variable that takes the value of 1 if the potential respondent is a member of top management team; 0 otherwise.
Mature MFI	Mature	A dummy variable that takes the value of 1 if the age of an MFI is over 8 years; 0 otherwise.
Offices		Number of offices that an MFI has.
Sustainable MFI	Sustainable	A dummy variable that takes the value of 1 when the sustainability (representing the ability to cover all costs) of an MFI is more than 100 percent; 0 otherwise.
Regulated		A dummy variable that takes the value of 1 if an MFI is regulated by some supervisory authority; 0 otherwise.
Profit margin		Operating Income divided by the Financial Revenue.
Percentage of female borrowers	Female percentage	The number of female active borrowers divided by the total number of active borrowers.
Outreach		A dummy variable that takes the value of 1 if the total number of clients that an MFI is serving is over 30,000; 0 otherwise.
Scale		A dummy variable that takes the value of 1 if an MFI size is large; 0 otherwise. Institutional scale is measured by the size of an institution's loan portfolio in US Dollars (USD). The measure of scale is regionalized to reflect differences in income levels across regions.

Continues on next page

Continued from last page

Variable	Abbrevation	Definition
Target market		A dummy variable that takes the value of 1 if the target market of an MFI is broad/high end; 0 otherwise. MFIs are classified on basis the average balance of loans served into broad/high end or low end. For international comparison, this balance is stated as a percentage of local income levels (GNI per capita).
For profit MFI	For profit	A dummy variable that takes the value of 1 if an MFI is a for-profit organization; 0 otherwise. MFIs are classified to be profit-oriented MFIs when profit generation is their main goal.
Africa		A dummy variable that takes the value of 1 if an MFI is located in Africa; 0 otherwise.
East Asia and The Pacific	Asia Pacific	A dummy variable that takes the value of 1 if an MFI is located in East Asia or the Pacific; 0 otherwise.
Eastern Europe and Central Asia	Europe Asia	A dummy variable that takes the value of 1 if an MFI is located in Eastern Europe or Central Asia; 0 otherwise.
Latin America and The Caribbean	America Caribbean	A dummy variable that takes the value of 1 if an MFI is located in Latin America or the Caribbean; 0 otherwise.
Middle East and North Africa	Middle East	A dummy variable that takes the value of 1 if an MFI is located in the Middle East or North Africa; 0 otherwise.
South Asia	South Asia	A dummy variable that takes the value of 1 if an MFI is located in South Asia; 0 otherwise.
Variables from miscellaneous sources		
OIC Member	OIC	A dummy variable that takes the value of 1 if a potential respondent is a staff member at an MFI is operating in an OIC member state; 0 otherwise. Source: http://www.oic-oci.org/states/?lan=en.
Collectivism—Individualism	Individualism	Individualism scores countries on a scale of 0 to 100 reflecting the degree to which individuals in a country consider themselves as autonomous personalities (higher score) or as members of close communities (lower score). Source: Begelsdijk and Welzel 2018.
Distrust—Trust	Trust	Trust scores countries on a scale of 0 to 100 reflecting the degree to which individuals in a country feel comfortable in unstructured situations (higher score) or stressed and anxious in such situations (lower score). Source: Begelsdijk and Welzel 2018.
English as official language	English official	A dummy variable that takes the value of 1 if an MFI is operating in a state where English is the official language; 0 otherwise. Source: Ethnologue 2015 and World Factbook 2015. Source for the retrieved list: http://www.emmir.org/fileadmin/user_upload/admission/Countries_where_English_is_an_official_language.pdf.

Notes. The first, second and third columns of this table report, for some key variables used in the empirical analysis, the variable name, its abbreviated name and its definition/sources, respectively.

Table B2
Country scores for Individualism and Trust

Sr. No	Country	Individualism	Trust	Sr. No	Country	Individualism	Trust
1	Albania	34.2	27.8	35	Mexico	28.7	22.9
2	Argentina	36.2	15.9	36	Moldova	27.2	21.3
3	Armenia	26.3	17	37	Montenegro	33.6	29.4
4	Azerbaijan	24.1	49.4	38	Morocco	5.3	33.8
5	Bangladesh	4.2	61	39	Nigeria	8.8	32.6
6	Belarus	42.1	33.8	40	Pakistan	9.5	26.6
7	Bosnia and Herzegovina	36.2	33.7	41	Palestine	–	25.6
8	Brazil	25.3	18	42	Peru	19.1	0
9	Bulgaria	46.6	18.9	43	Philippines	22.5	45.8
10	Burkina Faso	15.5	25.8	44	Poland	29.7	27.3
11	Chile	24.5	26.1	45	Portugal	40.8	31.2
12	China	29.6	78.9	46	Romania	34.7	17.3
13	Colombia	16.6	16.1	47	Russia	39.2	23.6
14	Croatia	43.7	17.9	48	Rwanda	15.9	49.6
15	Dominican Republic	26.2	7.8	49	Saudi Arabia	12.5	–
16	Ecuador	14.1	15	50	Serbia	40.5	17.1
17	Egypt	2.8	52.4	51	Slovakia	47	28.8
18	El Salvador	–	15.7	52	South Africa	23.3	45.5
19	Ethiopia	21.6	26.9	53	Tanzania	9.1	55.8
20	Georgia	23.2	21.6	54	Thailand	20.8	48.3
21	Ghana	7.6	41.1	55	Trinidad and Tobago	14.8	16.9
22	Guatemala	12.3	–	56	Tunisia	4	14.2
23	Hungary	44.2	40.5	57	Turkey	19	41.3
24	India	22.3	48.3	58	Uganda	15.2	39.7
25	Indonesia	4.8	40.6	59	Ukraine	36.3	16.2
26	Iraq	2.7	27.1	60	United States	52.6	42
27	Jordan	0	49.8	61	Uruguay	48	34.3
28	Kazakhstan	22.5	45.8	62	Uzbekistan	17.9	85.7
29	Kosovo	11.9	55.7	63	Venezuela	12.4	12.2
30	Kyrgyzstan	15.3	29.2	64	Vietnam	17.9	100
31	Lebanon	30.3	22.4	65	Yemen	7.8	6.7
32	Macedonia	35.6	20	66	Zambia	25.2	31.3
33	Malaysia	15.5	57.6	67	Zimbabwe	11.4	37.3
34	Mali	21	42.8	-	-	-	-

Appendix C
Additional estimates for main outcome variables

Table C1
Survey waves overview—Details on response statistics by survey wave

	Survey waves									
	Initial email		1st reminder		2nd reminder		3rd reminder		4th reminder	
Response statistics	N	%	N	%	N	%	N	%	N	%
Complete	328	66.9	109	66.9	79	65.8	49	59.8	41	52.6
Partial or break-off with sufficient information	43	8.8	9	5.5	18	15.0	13	15.9	11	14.1
Break-off or partial with insufficient information	43	8.8	17	10.4	8	6.7	12	14.6	12	15.4
Logged on to survey, did not complete any items	76	15.5	28	17.2	15	12.5	8	9.8	14	17.9
Total	490	100	163	100	120	100	82	100	78	100

Notes. This table shows the response statistics for the initial and reminder survey invitation emails. *N* is the number of observations in each subgroup; % expresses this number as a percentage of the corresponding total for each column. Data for this table are from the authors' survey.

Table C2
Survey waves overview—Details on survey waves by response statistic

	Survey waves											
	Initial email		1st reminder		2nd reminder		3rd reminder		4th reminder		Total	
Response statistics	N	%	N	%	N	%	N	%	N	%	N	%
Complete	328	54.1	109	17.9	79	13.0	49	8.1	41	6.8	606	100
Partial or break-off with sufficient information	43	45.8	9	9.6	18	19.2	13	13.8	11	11.7	94	100
Break-off or partial with insufficient information	43	46.7	17	18.5	8	8.7	12	13.0	12	13.0	92	100
Logged on to survey, did not complete any items	76	53.9	28	19.9	15	10.6	8	5.7	14	9.9	141	100

Notes. This table shows the response statistics for the initial and reminder survey invitation emails. *N* is the number of observations in each subgroup; % expressed this number a percentage of the corresponding total for each row. Data for this table are from the authors' survey.

Table C3

Testing main hypotheses—Multi-level modeling

		Model 1: Direct treatment effects		Model 2: Moderation due to culture		Model 3: Additional moderation due to translation	
Variables	n	Contact (1)	Response (2)	Contact (3)	Response (4)	Contact (5)	Response (6)
AULP	2,564	0.098***	0.019**	0.117***	0.022*	0.063**	0.029*
OIC	1,780	-	-	0.105***	0.025	–0.022	–0.023
AULP*OIC	887	-	-	–0.053*	–0.006	0.038	–0.024
Transl	3,338	-	-	-	-	–0.008	0.007
AULP*Transl	1,669	-	-	-	-	0.093**	–0.013
Transl*OIC	1,390	-	-	-	-	0.166**	0.060
AULP*Transl*OIC	692	-	-	-	-	–0.139**	0.025
Controls	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	5,128	5,128	5,128	5,128	5,128	5,128	

This table reports the marginal effects computed at the sample means from the corresponding multi-level logistic regression models (Models 1-3) described in the main text. We report standard errors in parentheses. *AULP* is a dummy variable that takes the value of 1 if a potential respondent is (randomly) provided with the announcement of undefined lottery prizes in the subject line of the survey invitation emails; 0 otherwise. *Transl* is a dummy variable that takes the value of 1 if a potential respondent received a survey translated into an official language of their country; 0 otherwise (see Appendix A.3 for criteria). *OIC* is a dummy variable that takes the value of 1 if a potential respondent is a staff member at an MFI is operating in an OIC member state; 0 otherwise. *AULP*OIC, AULP*Transl, Transl*OIC*, and *AULP*Transl*OIC* are the interaction terms for respective variables. The variables *Target Market* (takes the value of 1 if the target market of an MFI is *broad/high end;* 0 otherwise) and *Mature* (takes the value of 1 if the age of an MFI is over 8 years; 0 otherwise) are used as control variables in all models. For each model, two sets of estimated marginal effects are displayed for the two binary dependent variables, "Contact" and "Response", defined respectively as: a dummy variable that takes the value of 1 if a potential respondent opened the survey invitation email (0 otherwise); and a dummy variable that takes the value of 1 if a potential respondent fully completed the survey (0 otherwise). See Table B1 for data sources.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix D
Supplementary material for data quality

This Appendix describes and reports the results of the analyses we undertake to check whether our treatment, AULP, leads to unwanted side effects in terms of data quality. We explore this possibility in terms of non-response bias (other than for OIC countries, which is one of the main hypotheses) as well as for several other measures of data quality, such as missing items and speeding.

D.1 Data Quality—Non-response bias

We explore the risk of non-response bias (e.g. Groves & Peytcheva, 2008) using a logistic model, Model 6 (see Subsection 3.8 in the main paper). For this analysis, the dependent variable "NR", is a binary variable, equal to 1 for category 2.0 (eligible, "non-interview") and 0 for category 1.0 (returned questionnaire) of the AAPOR 2016 disposition codes. Thus, it indicates whether a particular email recipient was a non-respondent (NR = 1), as defined by AAPOR. Recall from Subsection 3.8 that we can check for the possibility of non-response bias for the non-AULP group by testing if the marginal effect of a particular characteristic X is statistically significant (i.e., individuals with that characteristic [e.g., "male"] are more or less likely to not respond). For the AULP group, we can determine if the AULP treatment itself affects non-response bias by testing if the marginal effect on the interaction term between X and the AULP treatment variable is statistically significant. This marginal effect is of primary interest, because we are mainly interested in, for example, whether more men do not respond (cf. control group) when they receive the AULP treatment.

We estimate Model 6 for 19 such X variables, which are the ones we used for the balance test. Thus, we run 19 regressions, one for each of the 19 Xs. For most X variables, the coefficient of the interaction term between X and the AULP treatment is statistically insignificant. This finding is reassuring; it suggests a low possibility of non-response bias resulting from the AULP treatment.

The coefficient of the interaction term between X and the AULP treatment turns out to be statistically significant for only 2 out of the 19 characteristics. In Table D1 we present the marginal effects from the logistic regressions that include these two variables, *English Official* and *American Caribbean*. Note that people who live in countries in which English is an official language are more likely to not respond when treated with AULP; those in countries situated in Latin America or the Caribbean are more likely to respond when treated with AULP. For these two cases then, the AULP treatment actually *alleviates* existing non-response bias, as indicated because the marginal effects of X and the interaction term between X and the AULP treatment take opposite signs. The estimation results for the full set of 19 observable characteristics appear in the Tables D2 and D3; the marginal effect of the interaction variable is not significant for the remaining 17 observable characteristics of the email recipients. Thus, the AULP treatment does not significantly influence non-response bias with respect to these relevant variables.

For the analyses in Tables D1, D2 and D3, we excluded category 3.0 (unknown eligibility, "non-interview") to define NR (leaving a sample of 2,190 observations); we also conduct a robustness check, using an alternative specification of non-response, in which NR takes the value of 1 for categories 2.0 (eligible, "non-interview") and 3.0 (unknown eligibility, "Non-Interview") and a value of 0 for category 1.0 (returned questionnaire) according to AAPOR 2016 disposition codes (in which case the estimation, uses all 5,128 observations). Our results remain unchanged for this alternative specification.[11] Therefore, we find little evidence that suggests that the AULP treatment might increase non-response bias—in terms of personal, MFI, or regional characteristics—in studies that adopt it to boost contact and response rates.[12]

D.2 Data quality—time variables, filled items ratio, and recontact

In this subsection, we present an analysis using Model 7 (see Subsection 3.8 of the main paper) for the data quality variables related to the time spent completing the survey, the number of filled versus missing questionnaire items, and recontact willingness. We first describe these variables and then present empirical estimates of Model 7. Table D7 provides precise definitions and summary statistics for the variables; Table D8 lists their Spearman correlations.

Time spent completing the survey . The time spent by respondents to complete a survey is commonly used as a proxy for data quality. For example, studies of speeding (e.g. Goeritz, 2006; Greszki, Meyer, & Schoen, 2015) explore whether a *shorter* time spent to complete a survey is associated with diminished data quality. The programming for our survey allowed respondents to pause and resume filling the questionnaire at any time (via the original link or the link in any of the reminder emails that were sent if the survey remained incomplete), re-accessing the same questionnaire, without any loss of their previously submitted an-

[11]These regression results are available on request.

[12]Note that the cooperation rates (see Subsection B.1 for the definition of Cooperation Rate 1 according to AAPOR 2016 standard definitions) therefore should not significantly differ between treatment and control (i.e., AULP and non-AULP) groups. We test this conjecture directly by regressing an indicator variable for cooperation on our treatment dummy (equivalent to Model 1 in Table 2 of the main paper) and fail to reject the hypothesis. We also check for heterogeneity across subgroups (i.e., estimate a version of Model 3, Table 2, using a cooperation indicator as dependent variable) but do not find evidence of heterogeneity. These regression results are available on request.

Table D1
Non-response analyses

Variables	(1)	(2)
AULP	0.054^{**}	0.000
	(0.024)	(0.023)
English Official	0.091^{***}	-
	(0.034)	-
AULP*English Official	-0.085^{*}	-
	(0.044)	-
America Caribbean	-	-0.142^{***}
	-	(0.038)
AULP*America Caribbean	-	0.137^{***}
	-	(0.049)
Observations	2,190	2,190
McFadden R2	0.004	0.006

Notes. This table reports the marginal effects computed at the sample means from the corresponding logistic regression model (Model 6). *AULP* is a dummy variable that takes a value of 1 if a potential respondent is (randomly) provided with an announcement of undefined lottery prizes in the subject line of the survey invitation emails and 0 otherwise. *English Official* in Column (1) is a dummy variable that takes a value of 1 if an MFI is operating in a state where English is the official language and 0 otherwise. *American Caribbean* in Column (2) is a dummy variable that takes a value of 1 if an MFI is located in Latin America or the Caribbean and 0 otherwise. The binary dependent variable NR takes a value of 1 for category 2.0 (eligible, "non-interview") and 0 for category 1.0 (returned questionnaire), according to AAPOR 2016 disposition codes. Standard errors clustered at MFI level are in parentheses. See Table B1 for data sources.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

swers.[13] Thus, we can unambiguously measure the total time a particular respondent took to complete a survey (i.e., time from first access to final submission), even if the respondent completed the survey over multiple visits.

Naturally, recording the time from first access to final submission produced some outliers (Subsection D.5, Figure D1, provides a box-plot of the time data), which we address in two ways: (1) transform the response time variable by taking the natural logarithm (*LnTime*) and (2) winsorize the upper 2.5 percentile values of the log-transformed response time (*WinLnTime*). In the empirical analysis, the time-related dependent variables use alternative (1). We display the results with the winsorized alternative (2) in Table D5. Furthermore, skip patterns in the questionnaire can influence the time spent completing the survey. To account for these patterns, we construct two additional measures by scaling *LnTime* and *WinLnTime* by the total number of items a respondent needed to answer to complete the survey (see Subsection D.3 for additional information), that is, *LnTime/Total Items* and *WinLnTime/Total Items*.

Number of filled versus missing questionnaire items. Nonrandomly missing data present a problem for statistical analysis (Wooldridge, 2010), so we need to check for any systematic patterns of missing data in replies to our survey (e.g. Heerwegh, 2006; Janke, 2014; Sánchez-Fernández, Muñoz-Leiva, & Montoro-Rios, 2012). We use the filled items ratio, or the ratio of answered to total questions/items in the survey (*Filled items ratio*); Subsection D5 further details the calculation of this ratio. The value of this ratio ranges from 0 to 1, where 0 corresponds to logging into the survey without filling any items and 1 corresponds

[13] Note that this feature also allowed us to avoid duplicate/multiple responses by the same respondent, as mentioned in Subsection 3.3 of the main paper.

Table D2
Non-response—Individual and regional characteristics

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
AULP	0.052	0.026	0.054**	0.042*	0.035*	0.031	0.000	0.027	0.031
	(0.042)	(0.030)	(0.024)	(0.024)	(0.021)	(0.022)	(0.023)	(0.021)	(0.023)
X	−0.057*	−0.059*	0.091***	0.043	0.069	0.060	−0.142***	−0.129*	0.048
	(0.034)	(0.030)	(0.034)	(0.033)	(0.054)	(0.048)	(0.038)	(0.066)	(0.037)
AULP* X	−0.030	0.004	−0.085*	−0.051	−0.093	−0.030	0.137***	−0.023	−0.011
	(0.048)	(0.041)	(0.044)	(0.044)	(0.072)	(0.061)	(0.049)	(0.086)	(0.049)
Observations	2,190	2,190	2,190	2,190	2,190	2,190	2,190	2,190	2,190
McFadden R2	0.005	0.004	0.004	0.001	0.001	0.002	0.006	0.004	0.002

Notes. This table reports the marginal effects computed at the sample means from the corresponding logistic regression model (Model 6) described in the main text. *AULP* is a dummy variable that takes the value of 1 if a potential respondent is (randomly) provided with the announcement of undefined lottery prizes in the subject line of the survey invitation emails; 0 otherwise. *X* represents a different variable depending on the column: *Male* for Column (1), *Top management* for Column (2), *English Official* for Column (3), *Africa* for Column (4), Asia Pacific for Column (5), Europe Asia for Column (6), America Caribbean for Column (7), Middle East for Column (8), and South Asia for Column (9). AULP*X is the interaction of the respective variable X with AULP for each regression. Definitions of these variables are in Appendix Table B1. The binary dependent variable *NR* takes a value of 1 for category 2.0 (eligible, "non-interview") and 0 for category 1.0 (returned questionnaire) according to AAPOR 2016 disposition codes. Standard errors clustered at MFI level are shown in parentheses. See Table B1 for data sources.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D3
Non-response—MFI characteristics

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
AULP	0.066**	0.0323	−0.006	0.046	0.028	0.049	0.029	0.049*	0.025	0.023
	(0.032)	(0.0209)	(0.030)	(0.033)	(0.020)	(0.031)	(0.024)	(0.026)	(0.026)	(0.026)
X	0.003	0.0001	−0.071**	0.071**	−0.038	0.015	−0.051	−0.017	−0.035	0.055*
	(0.031)	(0.0001)	(0.031)	(0.031)	(0.030)	(0.043)	(0.035)	(0.032)	(0.032)	(0.032)
AULP* X	−0.062	−0.0001	0.058	−0.029	0.009	−0.053	−0.011	−0.056	0.010	0.011
	(0.041)	(0.0001)	(0.041)	(0.041)	(0.033)	(0.057)	(0.045)	(0.041)	(0.041)	(0.041)
Observations	2,190	2,190	2,190	2,190	2,190	2,190	2,190	2,190	2,190	2,190
McFadden R2	0.002	0.001	0.003	0.004	0.002	0.001	0.003	0.003	0.002	0.004

Notes. This table reports the marginal effects computed at the sample means from the corresponding logistic regression model (Model 6) described in the main text. *AULP* is a dummy variable that takes the value of 1 if a potential respondent is (randomly) provided with the announcement of undefined lottery prizes in the subject line of the survey invitation emails; 0 otherwise. *X* represents a different variable depending on the column: *Mature* for Column (1), *Offices* for Column (2), *Sustainable* for Column (3), *Regulated* for Column (4), *Profit margin* for Column (5), *Female percentage* for Column (6), *Outreach* for Column (7), *Scale* for Column (8), *Target market* for Column (9) and *For profit* for Column (10). AULP*X is the interaction of the respective variable X with AULP for each regression. Definitions of these variables are Appendix Table B1. The binary dependent variable *NR* takes a value of 1 for category 2.0 (eligible, "non-interview") and 0 for category 1.0 (returned questionnaire) according to AAPOR 2016 disposition codes. Standard errors clustered at MFI level are shown in parentheses. See Table B1 for data sources.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

to a fully completed survey. The average values of the filled items ratio for those who partially filled the questionnaire with sufficient information, and those who partially filled the questionnaire with insufficient information, are 0.74 and 0.12 respectively.

Willingness to be contacted again in the future. We include the stated willingness of respondents to be contacted again *(Recontact)* as a data quality measure because, arguably, respondents who agree to be re-contacted are also more likely to have filled in the data accurately and/or are interested in the success of the study, as indicated by their willingness to invest further time in the research effort.[14] Further details about the variable calculation are in Subsection D5.

[14] Only 40 people actively opted out of receiving future survey invitations.

Data quality results. Table D4 contains the results we obtained from estimating Model 7 for four (*LnTime*, *LnTime/Total Items*, *Filled items ratio, Recontact*) of the six dependent variables; Table D5 provides the results that account for time outliers using the winsorized versions of the time variables (*WinLnTime* and *WinLnTime/Total Items*). Models 7a–7c, which involve continuous dependent variables (*LnTime*, *LnTime/Total Items, Filled items ratio*), are estimated using OLS, but we use a logistic regression for Model 7d, which involves the binary dependent variable *Recontact*. For Model 7d, Table D4 reports marginal effects for each independent variable, calculated at the means of the other independent variables. Note that the sample sizes for the estimation differ across models (see the Table D4 notes for details).

As mentioned in Subsection 3.8 of the main paper, this analysis aims to check for any clear patterns of differences in the data quality measures between the treatment and control groups, as well as the effects of the moderating variables. Table D4 shows that the AULP variable is not statistically significant in any of the regressions, by itself or in interaction. Thus, characteristics of data quality do not significantly differ between the treatment and control groups. Only the provision of translations is significantly negatively correlated (at the 10% level) with the likelihood that a respondent is willing to be contacted again. The results for the time variables (Models 7a and 7b) do not change when we use the two alternative winsorized measures either (Table D5).

D.4 Additional data quality measures - answer patterns, break-offs and straightlining

We further tested whether responses to individual questions and sub-questions (i.e., items) in the questionnaire differ across treatment and control groups but find no significant differences, with the sole exception of a question that asked about the use of interest-based financial products. More respondents reported being from MFIs that offered interest-based products in the treatment group, which corresponds with one of our main findings, namely, that the intervention is more effective for non-OIC countries than OIC countries (note that interest-based financial products are likely to be less prevalent in OIC countries, in line with Islamic strictures regarding interest)[15].

Break-offs and straightlining affected only a very small part of the sample. For break-offs, Table 1 shows, with regard to the main outcome rates, that break-offs and partials together affect only 186 of the 5,218 (3.56%) respondents in our sample. Among those, break-off with insufficient information represents only 1.76%, (rounded to 1.8%) and break-off with sufficient information represents the remainder (1.80%). With regard to straightlining, only two respondents answered in a straight line for one of the compulsory questions, and neither of them had received the AULP treatment. Further details on break-offs and straightlining are provided in Subsection D5.

Therefore, across all these additional analyses, the findings suggest that the AULP treatment does not affect data quality negatively.

D.5 Descriptive details of data quality

This section presents additional material, such as definitions and descriptive statistics, related to the data quality analysis in the main body of the paper, specifically concerning time spent completing the survey, missing data, recontact, and straightlining.

Time spent completing the survey. We present below a box plot for the time spent to fully complete a survey (recall that we use Response Rate 1 with fully completed surveys in the numerator throughout our paper and that we use the time spent to fully complete a survey in the calculation of our *LnTime* measure for the data quality analysis).

The box plot shows that the median time it took respondents to fully complete a survey was slightly higher than the 5 minutes we announced the survey would take in the subject lines of our survey invitation emails. However, respondents were allowed to pause and resume filling the questionnaire, with the link in the original survey invitation email, as well as the link included in each reminder email, always leading back to the same survey with any potential prior responses saved. The time recorded captures the time a respondent first accessed the survey till final submission. Recording time in this way naturally led to a few outliers, which can be discerned in the box plot. We dealt with these outliers by taking natural logs of the time variable as well as winsorizing (see main paper text for details).

Missing data. While our response measure only takes fully completed surveys into account, we utilize information about missing data for determining the overall outcome rates for our survey (see Table 1 in the main paper); and for the calculation of our filled items ratio which we employed in our data quality analysis (see Table D7). The following paragraphs provide supplementary information to the discussion of missing data.

Outcome rates. A description of break-offs is provided in the paper using Table 1. Overall, 1.8 percent of the total eligible respondents (i.e., of the full contact list with unique email addresses) partially filled the questionnaire with sufficient information; 1.8 percent of the total eligible respondents partially filled the questionnaire with insufficient information; and 2.8 percent of the total eligible respondents only logged on to the survey but did not provide any information.

[15] We do not report the actual regressions here for brevity, but the results are available on request.

Table D4
Data quality

Variables	Model 7a: Log of time (LnTime) (1)	Model 7b: Log of time to total items ratio (LnTime/total items) (2)	Model 7c: Filled items ratio (3)	Model 7d: Recontact (4)
AULP	0.129	0.017	–0.005	0.056
	(0.298)	(0.030)	(0.057)	(0.076)
OIC	–0.126	–0.018	–0.057	–0.048
	(0.305)	(0.030)	(0.107)	(0.130)
AULP*OIC	–0.179	–0.019	–0.040	0.227
	(0.432)	(0.044)	(0.138)	(0.233)
Transl	0.218	0.006	0.056	-0.115^{*}
	(0.296)	(0.027)	(0.052)	(0.062)
AULP*Transl	0.105	0.021	–0.060	–0.077
	(0.430)	(0.041)	(0.071)	(0.090)
Transl*OIC	0.536	0.054	0.117	0.124
	(0.443)	(0.040)	(0.115)	(0.137)
AULP*Transl*OIC	0.200	0.009	0.064	–0.251
	(0.621)	(0.059)	(0.151)	(0.242)
Controls	Yes	Yes	Yes	Yes
Constant	1.989^{***}	0.192^{***}	0.662^{***}	-
	(0.218)	(0.022)	(0.047)	-
Observations	606	606	933	724
(McFadden) R2	0.024	0.020	0.022	0.025
Observations per group				-
n(AULP=1)	329	329	514	387
n(OIC=1)	246	246	362	291
n(AULP=1,OIC=1)	130	130	194	152
n(Transl=1)	413	413	637	518
n(AULP=1,Transl=1)	222	222	350	276
n(Transl=1,OIC=1)	211	211	302	255
n(AULP=1,Transl=1,OIC=1)	112	112	162	134

Notes. This table reports the OLS regression coefficients (Columns 1-3) and the marginal effects computed at the sample means from a logistic regression (Column 4) for Model 7. The dependent variables for each column are as follows: Column 1 uses *LnTime* (log value of the time spent to complete the survey); Column 2 uses *LnTime/total items* (ratio of the time taken to complete the survey with respect to the total items in the questionnaire); Column 3 uses the *Filled items ratio* (number of answered questions/items in the survey divided by the total number of questions/items); and Column 4 uses *Recontact* (dummy variable equal to 1 if the respondent is willing to be contacted again). The models in Columns 1 and 2 are estimated using the sample of respondents who fully complete the questionnaire; that in Column 3 is estimated using the sample of respondents who partially/fully complete the survey, along with those who logged into the survey but did not fill in any items; and the model in Column 4 is estimated using the sample of respondents who partially/fully completed the survey. *AULP* is a dummy variable that takes a value of 1 if a potential respondent is (randomly) provided with an announcement of undefined lottery prizes in the subject line of the survey invitation emails and 0 otherwise. *Transl* is a dummy variable that takes a value of 1 if a potential respondent received a survey translated into an official language of their country and 0 otherwise (Appendix A0.2 for criteria). *OIC* is a dummy variable that takes a value of 1 if a potential respondent is a staff member at an MFI is operating in an OIC member state and 0 otherwise. *AULP*OIC, AULP*Transl, Transl*OIC,* and *AULP*Transl*OIC* are the interaction terms for the respective variables. *Target Market* (= 1 if the target market of an MFI is *broad/high end* and 0 otherwise) and *Mature* (=1 if the MFI is older than 8 years and 0 otherwise) are control variables in all models. Robust standard errors are shown in parentheses. See Table B1 for data sources.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D5
Model 7 with winsorized time variables

Variables	(1) 7a: Winsorized log of time (wLnTime)	(2) 7b: Winsorized log of time to total items ratio (wLnTime/total items) to total items ratio (wLnTime/total items)
AULP	0.125	0.017
$n(AULP = 1) = 329$	(0.293)	(0.029)
OIC	−0.033	−0.007
$n(OIC = 1) = 246$	(0.276)	(0.027)
AULP*OIC	−0.240	−0.026
$n(AULP = 1, OIC = 1) = 130$	(0.421)	(0.042)
Transl	0.188	0.004
$n(Transl = 1) = 413$	(0.293)	(0.026)
AULP*Transl	0.129	0.023
$n(AULP = 1, Transl = 1) = 222$	(0.420)	(0.040)
Transl*OIC	0.473	0.047
$n(Transl = 1, OIC = 1) = 211$	(0.420)	(0.038)
AULP*Transl*OIC	0.229	0.013
$n(AULP = 1, Transl = 1, OIC = 1) = 112$	(0.605)	(0.058)
Controls	Yes	Yes
Constant	2.014^{***}	0.192^{***}
	(0.193)	(0.019)
Observations	606	606
R-squared	0.023	0.018

Notes. This table reports the results of the OLS regressions (Model 7 in the paper) for those who fully complete the questionnaire. *wLnTime* represents the winsorized values of log of time spent to complete the survey; *wLnTime/total items* is a ratio representing the winsorized log of time taken to complete the survey with respect to the total items in the questionnaire. *AULP* is a dummy variable that takes the value of 1 if a potential respondent is (randomly) provided with the announcement of undefined lottery prizes in the subject line of the survey invitation emails; 0 otherwise. *Transl* is a dummy variable that takes the value of 1 if a potential respondent received a survey translated into an official language of their country; 0 otherwise (see Appendix A.2 for criteria). *OIC* is a dummy variable that takes the value of 1 if a potential respondent is a staff member at an MFI is operating in an OIC member state; 0 otherwise. *AULP*OIC, AULP*Transl, Transl*OIC,* and *AULP*Transl*OIC* are the interaction terms for respective variables. The variables *Target Market* (takes the value of 1 if the target market of an MFI is *broad/high end;* 0 otherwise) and *Mature* (takes the value of 1 if the age of an MFI is over 8 years; 0 otherwise) are used as control variables in all models. Standard errors clustered at MFI level are shown in parentheses. See Table B1 for data sources.
$^{***}\ p < 0.01$, $^{**}\ p < 0.05$, $^{*}\ p < 0.1$.

To provide finer detail, we report here the above percentages for the AULP and no-AULP groups separately. For the AULP and no-AULP groups, 1.75 and 1.91 percent of the total eligible respondents in the respective groups partially filled the questionnaire with sufficient information; 2.15 and 1.44 percent of the total eligible respondents in the respective groups partially filled the questionnaire with insufficient information; and 3.32 and 2.18 percent of the total eligible respondents in the respective groups only logged on to the survey but did not provide any information.

To clarify the meaning of these numbers in the context of our survey, we provide a more granular overview for break-offs and partials with respect to question numbers of our survey questionnaire in the table below. The median for “break-off with insufficient information” is Question number 1; and the median for “break-off with sufficient information” is Question number 3. The same question number can be deemed sufficient or insufficient, as respondents were asked 5, 7 or 10 questions based on their responses.

Filled items ratio. Based on the answers of the MFI practitioners, they were asked 5, 7 or 10 questions, as mentioned above. In addition, since some questions had subquestions, we define our filled items ratio as the ratio of the number of items (i.e., sub-questions) that a respondent answered, divided by the total number of items in the survey that the respondent would have been asked (i.e., total pos-

Table D6
Overview for break-offs and partials with respect to question numbers

Question Number	Break-off or partial with insufficient information		Partial or breakoff with sufficient information	
	N	%	N	%
1	74	80.43	–	–
2	6	6.52	–	–
3	11	11.95	39	41.48
4	1	1.09	30	31.92
5	–	–	8	8.51
6	–	–	2	2.13
7	–	–	9	9.57
8	–	–	3	3.19
9	–	–	0	–
10	–	–	3	3.19
Total	92	-	94	-

Table D7
Variables used in data quality checks

Variable name	Definition	N	Mean	St. Dev.
Non-response	Binary variable taking a value of 1 for category 2.0 (eligible, "non-interview") and 0 for category 1.0 (returned questionnaire) according to AAPOR 2016 disposition codes.	2190	0.68	0.47
Non-response—alternative specification	Alternative specification of non-response is a binary variable taking the value of 1 for categories 2.0 (eligible, "non-interview") and 3.0 (unknown eligibility, "Non-Interview"), and 0 for category 1.0 (returned questionnaire) according to AAPOR 2016 disposition codes.	5128	0.86	0.34
Log of time spent completing the survey (LnTime)	The survey system records the time duration for completing the survey in minutes. LnTime represents the log value of the time spent.	606	2.41	2.13
Log of time to total items ratio (LnTime/total items)	This ratio represents the time taken to complete the survey with respect to the total items in the questionnaire.	606	0.22	0.19
Winsorized log of time spent completing the survey (wLnTime)	Winsorized version of Lntime. As LnTime is positively skewed, we use one tail winsorization (top 2.5 percent).	606	2.39	2.07
Winsorized log of time to total items ratio (wLnTime/total items)	Ratio of wLnTime and the total items in the questionnaire	606	0.22	0.19
Filled items ratio	Number of answered questions/items in the survey divided by the total number of questions/items. The values range from 0 to 1, where 0 corresponds to logging into the survey without filling any items and 1 corresponds to fully completed surveys.	933	0.74	0.41
Willingness to be contacted again (Recontact)	Respondents were asked to indicate whether they are willing to be contacted again in case of any query, conditional upon getting to the last page of the survey. Recontact=1 if they are willing to be contacted again.	724	0.76	0.43

Note. This table reports variable definitions, number of observations (N), means and standard deviations (SD) of the dependent variables used for the data quality estimates. LnTime, LnTime/total items, wLnTime and wLnTime/total are determined for those who completed the survey; Filled items ratio is determined for those who partially/fully completed the survey and those who logged into the survey, but did not fill in any items; Recontact is missing for those who did not fill in any survey questions. The data are from the authors' survey.

Table D8
Spearman correlations

	AULP	OIC	Transl	LnTime	LnTime/total-items	Filled items ratio	Recontact
AULP	1	-	-	-	-	-	-
	N = 724	-	-	-	-	-	-
OIC	-0.020	1	-	-	-	-	-
	N = 724	N = 724	-	-	-	-	-
	p = 0.590	-	-	-	-	-	-
Transl	-0.005	0.292***	1	-	-	-	-
	N = 724	N = 724	N = 724	-	-	-	-
	p = 0.883	p<0.000	-	-	-	-	-
LnTime	0.051	0.168***	0.144***	1	-	-	-
	N = 606	N = 606	N = 606	N = 606	-	-	-
	p = 0.208	p<0.000	p = 0.0004	-	-	-	-
LnTime/total items	0.055	0.116***	0.114***	0.928***	1	-	-
	N = 606	N = 606	N = 606	N = 606	N = 606	-	-
	p = 0.175	p = 0.004	p = 0.005	p<0.000	-	-	-
Filled items ratio[a]	0.035	0.017	-0.163***	-	-	1	-
	N = 724	N = 724	N = 724	-	-	N = 933	-
	p = 0.349	p = 0.653	p<0.000	-	-	-	-
Recontact	-0.009	0.029	-0.127***	0.053	0.028	0.755***	1
	N = 724	N = 724	N = 724	N = 606	N = 606	N = 724	N = 724
	p = 0.800	p = 0.443	p = 0.001	p = 0.192	p = 0.487	p<0.000	-

Note. This table reports Spearman correlations between the variables used in the data quality estimates. N is the number of observations. *AULP* is a dummy variable that takes the value of 1 if a potential respondent is (randomly) provided with the announcement of undefined lottery prizes in the subject line of the survey invitation emails; 0 otherwise. *Transl* is a dummy variable that takes the value of 1 if a potential respondent received a survey translated into an official language of their country; 0 otherwise (see Appendix A.2 for criteria). *OIC* is a dummy variable that takes the value of 1 if a potential respondent is a staff member at an MFI is operating in an OIC member state. *LnTime* represents the log value of the time spent to complete the survey; *LnTime/total items* is a ratio representing the time taken to complete the survey with respect to the total items in the questionnaire; *Filled items ratio* (which is the number of answered questions/items in the survey divided by the total number of questions/items) and *Recontact* is a dummy variable taking the value of 1 when the respondent is willing to be contacted again. The data stems are from the authors' survey.
[a] No correlation reported between Lntime (LnTime/total items) and Filled items ratio because LnTime is available for only for those who completed the survey, whereas Filled items ratio is 1 for all those who completed the survey.
*** p<0.001.

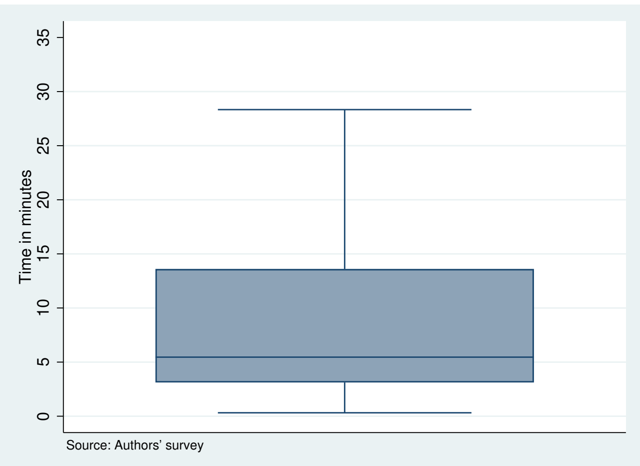

Figure D1. Time spent completing the survey

sible items). A question with no sub-questions was treated as one item. The average values of the filled items ratio for those who partially filled the questionnaire with sufficient information, and those who partially filled the questionnaire with insufficient information, are 0.74% and 0.12%, respectively.

Recontact. In total, 724 out of 792 respondents, who fully or partially completed the survey, provided an answer to the question about their willingness to be re-contacted in the future. Out of those, 75.83% (549 respondents total) expressed willingness to be re-contacted. The following details provide support for our choice of using Recontact as a data quality measure: Out of the 606 respondents, who fully completed the survey, 90.09% (546 respondents total) indicated willingness to be contacted again. All 94 respondents, who returned the survey with sufficient information, provided an answer to the question about their willingness to be re-contacted in the future. 96.8% of those respondents (i.e., 91 respondents total) expressed willingness to be re-contacted. In comparison, only 24 out of the 92 respondents, who returned the survey with insufficient information, provided an answer to the question about their willingness to be re-contacted in the future. Not a single one of them expressed willingness to be contacted again.

Straight lining. In our questionnaire, only the questions numbered 5, 7 and 8 were prone to straight lining because: a) they had sub-questions; and b) they had a grid format with a range of answer choice from zero to 100. Question 8 was compulsory for all respondents. Questions 5 and 7 on the other hand were not compulsory but were both presented to some of them depending on their earlier response to a mandatory question (specifically, this depended on the type of their institution). Of our respondents, only approximately 7 percent were asked to respond to all three questions 5, 7, and 8. We found that in our dataset, only two respondents, who were not presented with questions 5 and 7, answered in a straight line for the compulsory question 8. Neither of them had received the AULP treatment, i.e., both were in the control group.