



Engaging academic staff in the quality assurance system in higher education: A field experiment

Miguel A. Meléndez-Jiménez, Antonio J. Morales^{*}, Javier Rodero Cosano

Departamento de Teoría e Historia Económica, Universidad de Málaga, Plaza El Ejido s/n, Málaga 29013, Spain

ARTICLE INFO

JEL codes:

C93
D91

Keywords:

Field experiment
Behavioural intervention
Quality assurance programme
Faculty satisfaction survey
Higher education
Personalised message
Nudging
Norm-based
Moral suasion
Moral duty

ABSTRACT

A randomised field experiment was conducted at the University of Malaga in 2017 to examine personalised norm-based messages' effects on response rates to the annual faculty satisfaction survey. This research had a factorial design that combined formal versus informal writing styles with four norm-based messages: no norm, moral suasion, moral duty and social norm. The results reveal that two treatments – informal-moral duty and informal-social norm – both had the strongest positive effect. The university's quality assurance unit applied the informal-moral duty approach the following academic year, thereby increasing the overall response rate to the annual faculty satisfaction survey by 20% over the pre-intervention year.

1. Introduction

Employee engagement within companies is widely thought to be pivotal to businesses' success (Harter et al., 2002; Robertson-Smith, 2009; Sundaray, 2011; Bedakar & Pandita, 2014). Organisations can achieve higher productivity and better performance by fostering a culture that enables staff members to engage more fully in their work (Baumruk, 2006). Of the seven commonly accepted drivers of worker engagement (Robertson-Smith, 2009), the present study focused on how features of the two-way communication system between workers and their organisation affect employee engagement. The investigation took place at the University of Malaga and involved a field experiment that sought to increase faculty members' engagement with the university's quality assurance programme.

The Bologna Declaration of 1999 concentrated on strengthened European universities' quality assurance systems (European Education and Culture Executive Agency (Eurydice) 2020). Each university subsequently created a quality assurance unit (QAU) to ensure that stakeholders' (i.e. students and academic and administrative staff) opinions about how well university services and facilities function, among other issues, are taken into account. A report on how higher education

institutions' quality assurance process has evolved in Europe since the Bologna Declaration asserts that student and staff participation is a key principle in the development of strong quality assurance systems (Loukkola & Zhang, 2010). Many forms of participation exist, ranging from formal involvement in governing or consultation bodies to the provision of information through regular surveys (e.g. at the end of each year and course).

Loukkola & Zhang (2010) further report that academic staff's engagement level is relatively high, with a maximum of 90.5% found for formal involvement in self-evaluation activities. However, the survey response rate shows the lowest level of participation at 53.6%. At the University of Malaga, for example, the QAU conducts annual faculty satisfaction surveys at the end of each academic year, which has had similar low response rates (i.e. typically around 50%).

This weak participation in surveys poses a significant problem because, to develop effective policies focused on improving academic institutions' services and functions, administrations must collect information directly from those who participate in academic programmes. This challenge highlights the importance of finding efficient survey designs that increase response rates, which in turn facilitate the gathering of more information and ensure fewer missing data, thereby

^{*} Corresponding author.

E-mail address: amorales@uma.es (A.J. Morales).

Table 1
Summary of experimental treatments and sample sizes (2,271).

Writing style	Norms			
	No norm	Moral suasion	Moral duty	Social norm
Formal	<i>n</i> =196	<i>n</i> =196	<i>n</i> =175	<i>n</i> =185
Informal	<i>n</i> =400	<i>n</i> =387	<i>n</i> =363	<i>n</i> =369
	male=202 female=198	male=191 female=196	male=194 female=169	male=176 female=193

2 × 4 factorial design. The Formal-No Norm treatment (grey-shaded) is the treatment implemented in the University of Malaga and it is our baseline treatment. Note that for the Formal treatments, due to a mistake in the implementation of the experiment, we can only deduce the male/female ratio, though we do not have the gender information of the participants.

ensuring a higher degree of representativeness regarding the target population.

In 2017, a large-scale field experiment was carried out to find a more efficient design that would increase response rates to the University of Malaga's faculty satisfaction surveys by incorporating behavioural effects. The intervention consisted of creating various versions of the base email that the QAU typically sends to the academic staff inviting them to participate in an online survey. Around 2,300 faculty members were randomly selected to receive different versions of this message. Using a 'natural' field experiment methodology, academics were observed in their regular environment without being told that they were participating in an intervention (Harrison & List, 2004).

The current study had a factorial design. One treatment dimension was to provide information on norms in the email. According to Bicchieri & Dimant (2019), 'a norm is a behavioural rule that prescribes a certain behaviour to a specific group of people in a specific class of situations.' For each norm, a small paragraph describing the norm was added to the base email just before the line on which faculty members were given the link to the online survey.

Three types of norms were included. The first was moral suasion, which has been found to be an effective way to encourage cooperation in public good games (Dal Bó & Dal Bó, 2014). The second norm was moral duty, which has been shown to strengthen voluntary tax compliance (Blumenthal et al., 2001), while the last was social norm, which has been found to improve energy conservation (Allcott, 2011). An extensive body of literature provides compelling evidence of norms' positive effects on desired behaviours. Some of these prior studies are reviewed in section two.

The present research's second treatment dimension was the letter's style. The base email uses quite formal wording. Many previous studies involving interventions conducted by the Behavioural Insights Team (2012, 2013, 2014) have shown that one effective way to increase the response rate to letters is personalising messages, for instance, using the recipients' first name and a more informal style (Behavioural Insights Team, 2012). The current field experiment's results confirm that two informal treatments – informal-moral duty and informal-social norm – can yield gains of 15 percentage points over the baseline treatment.

Since faculty satisfaction questionnaires are administered annually, norm nudging's impact over various years could be analysed. The University of Malaga's QAU was so pleased with the intervention's results that, although the experiment ended after the findings were consolidated, the unit used the email based on the informal-moral duty approach to invite academic staff to participate in the next annual satisfaction survey. Faculty engagement that year further increased up to nearly 60%.

2. Behavioural intervention

The QAU conducts an annual faculty satisfaction survey at the end of each academic year before the summer break. The questionnaire is distributed online, so faculty members are invited to participate via an email message. The survey is available for five weeks, with reminders

sent out weekly. The present behavioural intervention took place in June 2017.

2.1. Intervention design

Table 1 provides a summary of the treatments administered during the experiment. As mentioned previously, we have a factorial design, with one dimension referring to the writing style (i.e. formal vs informal) and the other component exploring the norms' effects. The latter dimension included an extra line referring to three different norms (i.e. moral suasion, moral duty and social norm). Because the informal writing treatments included gendered nouns in Spanish (unlike the formal writing), we produced 12 email messages (four formal messages, four informal female messages and four informal male messages) and the QAU created 12 different subsamples (with a size of 202) from the total population of faculty members, 2,418. Table 1 displays the number of emails that reached recipients for each subsample.¹

The emails sent are available in Appendix A. They were all different versions of a base email that the QAU had sent in previous years to invite academic staff to participate in the annual faculty satisfaction survey. The latter email was written in a formal style using generic male past participle nouns (i.e. in Spanish). A formal opening salutation of 'Dear Sir/Madam' was followed by the first body paragraph of the letter, in which the recipients were informed that they had been invited to participate in the survey, and they were provided with a link to the questionnaire. The closing salutation thanked the recipients for their collaboration, and the letter was signed by the QAU.

The base email evidently relied on the academic staff's intrinsic motivation to participate in the survey. Many previous studies have shown how intrinsic motivation is a powerful driver of many activities, such as smoking cessation (Volpp et al., 2009), tax compliance (Dwenger et al., 2016), blood donation (Lacetera et al., 2012), charitable giving (Landry et al., 2010) or exercise (Charness & Gneezy, 2009).

The current experiment's treatments differed first in terms of the letter's writing style. The formal style was replaced by an informal version, with greetings (e.g. 'Hi Antonio') using the faculty member's first name. The letter was signed by 'Eva', the real first name of a QAU employee at the time of the intervention. In Spanish, the letter's informal version had to be gender sensitive, that is, the Spanish translation of the past participle 'invited' is 'invitado' for a male recipient and 'invitada' for a female.

Both formal and informal treatments were included to reflect the Behavioural Insights Team (2012, 2013, 2014) findings based on a large number of trials that examined ways to increase letter response rates. One key element is that behavioural interventions must attract attention, and an effective way of doing this is to personalise messages. For example, the Behavioural Insights Team found that adding someone's name to a generic text increases the amount of money paid (Behavioural Insights Team, 2013) and the collection of taxes owed (Behavioural Insights Team, 2012). In addition, personalised emails, as opposed to generic ones, significantly increase the number of people who decide to give one day's wages to a charitable cause (Behavioural Insights Team, 2014).

In addition, the present research's treatments varied in whether a short paragraph or sentence was added to the email's other content. This extra paragraph or sentence was placed just before the line on which faculty members were provided the link to the online survey, and the

¹ At the time of the current field experiment, the QAU's database covered 2,418 faculty members, 923 of whom were female (38.17%). The total number of invitation emails that reached faculty members was 2,271, 849 of whom were female (37.38%). To protect the respondents' privacy, the research team did not participate in the process of allocating faculty members to treatments/messages. The team was solely in charge of designing the intervention and composing the 12 different versions of the invitation emails and reminders.

added sentence gave information on a norm. This variation reflected the extensive body of literature offering evidence of how this kind of information is a powerful tool with which to change recipients' behaviour (Thaler & Sunstein, 2008; Sunstein, 2019).

In the current experiment, three different short paragraphs or sentences were added one at a time to the base letter. The first one focused on moral suasion: 'We need your help to improve the University of Malaga. If more of us complete the questionnaire, we will have more opportunities to make improvements.' This message was based on an institutional benefits argument for participating in the survey, which appealed to respondents' willingness to cooperate with the institution for the greater good. Much evidence has been found that urging recipients to do the right thing has a positive effect on cooperation (Ferraro et al., 2011; Dal Bó & Dal Bó, 2014). This version also contained an explicit reference to the University of Malaga in order to appeal to a sense of group identity and the need for a joint effort (i.e. 'more of us') to ensure more effective improvements could be made. The existing literature links group identity with social preferences that encourage individuals to take decisions for everyone's greater good.

The second addition was based on the related concept of moral duty: 'One of your duties as a faculty member is to help the University of Malaga improve.' In this case, the goal was to influence faculty members by increasing the costs associated with deviations from a moral standard. A wide range of studies have highlighted the role of moral obligation in explaining individual behaviour, for instance, tax compliance (Blumenthal et al., 2001; Alm & Torgler, 2011; Bott et al., 2019) or the Dictator Game (Capraro & Vanzo, 2019). An additional source of moral obligation is reciprocity (Fehr & Gächter, 1998), in which faculty members are expected to deliver a number of things in exchange for their wages. This second message was a reminder that one of the academic staff's duties is participation in the survey.

The last variation focused on social norms: 'A majority of the academic staff at the University of Malaga completes this survey.' This sentence conveyed information about other employees' behaviour, serving as a nudge (Bicchieri & Dimant, 2019) whose mechanism of action relies on social norms in order to provide or elicit social expectations. One challenge when using social norms is that the message writers need to know the recipients' reference group in advance (Hallsworth et al., 2017; Bicchieri & Dimant, 2019).

In the present study, another option would have been to group the faculty by schools, but the response rates in previous years had varied heterogeneously across programmes. The average response rate was 51.7%, yet some schools did not reach 30% (i.e. medicine) while others achieved 80% (i.e. management). Thus, the decision was made to keep the reference group to a generic allusion to the university's academic faculty and include the appealing wording 'a majority of'. Many researchers have previously confirmed the impact of providing information about other people's behaviour with reference to the recipients' own behaviour regarding, for instance, environmental economics (Ito et al., 2018), tax collection (Hallsworth et al., 2017), charitable giving (Frey & Meier, 2004; Shang & Croson, 2009) or voter turnout (Gerber et al., 2008).

2.2. Procedures

The QAU was in charge of the implementation of the field experiment, which required the allocation of the 2,418 faculty members among the 12 subsamples. The QAU followed the following procedure. Given that 8 of the 12 subsamples were gendered ones (four all-female informal and four all-male informal samples), the QAU allocated roughly 202 members for each of the 8 informal subsamples, exhausting 1,812 faculty members (806 male and 806 female members). This left 806 faculty members (690 male and 116 female), who were evenly allocated among the 4 formal subsamples. Hence, each formal subsample was composed of 172 male and 29 female faculty members. This allocation procedure produces an unbalanced gender composition for

each of the formal samples (85.57% male and 14.42% female).² The QAU sorted the academic staff by gender, and then, the unit arranged the faculty within each gender group by increasing order of their Spanish Identity (ID) card number. For each gender, the unit started to allocate them to the eight samples.³

To invite academics to participate in the online survey, the QAU followed the same protocol of the previous year: an invitation email. The letters were sent on Monday 5 June 2017 from the unit's official email address, and the subject line read 'Invitation to participate in the faculty satisfaction survey'. Each academic staff member was provided a link to the online satisfaction survey with a personalised token. In this way, the QAU could send reminders only to those members who had not yet filled in the questionnaire, on every Monday until 10 July. The link with token allowed faculty members to save and restart the survey at the point where they left off, before finishing and sending their answers.⁴

The survey software (i.e. LimeSurvey) automatically sent the weekly reminders to the staff who had not yet completed the questionnaire. Reminders were also personalised or not according to the same intervention design as the invitation messages. In addition, the personnel from some schools' administration services sent general reminders about the survey to all their faculty members every week during the field experiment. These messages could not be controlled for because they were decentralised and the only people aware of the intervention were the QAU and research team. The reminders were sent to all faculty staff, as had been done in previous years.

The LimeSurvey software recorded each respondent's last block of fully answered questions and timestamped the questionnaires with the finish date. LimeSurvey also provided information about the number of respondents who clicked on the invitation link and the number of academics who finished all five blocks of questions. Even though survey participation and answers were anonymous, the first block asked about the respondents' current position at the university, tenure category and school. Assuming *truthful* revelation, this information allow us to obtain the rate of responses by university position and school for the different treatments (the amount of socio-demographic information requested was kept to a minimum to guarantee anonymity; in particular, gender is not asked).

² Because answers to the questionnaire were anonymous, for those participants in any formal treatment we do not have their gender information. For this reason, we cannot directly analyse the treatment effect of the writing style comparing the participation rates among our samples. However, since we do know the male/female ratio in the formal treatments, we can rely on bootstrapping methods to assess the existence of treatment effects associated to the writing style. See footnote 11 for a discussion of the use of bootstrap methods for the analysis of experimental data.

³ This allocation procedure was clearly not entirely random. Since 1990, ID numbers have been randomly allocated to Spanish citizens, but, prior to 1990, blocks of numbers were randomly distributed among the different Spanish provinces and individual ID numbers were assigned to citizens when they visited the issuing offices to obtain their Spanish ID cards. Nearly 99% of the faculty members in 2017 were Spanish as only 30 faculty members – 3 women and 17 men – were registered in the database with a document different from a Spanish ID card, and the staff turnover in 2017, 2018 and 2019 was almost zero. As in any other Spanish university, academic mobility is a rare event. To assess whether the ID number assignment approximated random allocation, the distribution of one socioeconomic variable (i.e. tenure category) was analysed across the different treatments (see Appendix C Table C1). In most treatments, the percentage of permanent (i.e. tenured) faculty was around the overall average, although this type of faculty was under-represented in one treatment. Subsection 3.4 discusses how this may have introduced bias into the results.

⁴ The questionnaire was quite long because the academic staff were asked to use a 5-point Likert-scale to evaluate the service quality of the university's central administration, school administrations and all the degree programmes in which respondents taught. In total, they had to fill in five blocks of questions.

Table 2
Sample sizes, gender distribution and choices by message.

Treatment	Sample Gender Composition	Size	% female	Response rates		
				Click	Block I	Completed
Formal – no norm	Mixed	196	14.3	0.566	0.474	0.459
Informal – no norm	Male sample	202	0	0.515	0.416	0.396
	Female sample	198	100	0.601	0.535	0.520
Formal – moral suasion	Mixed	196	14.3	0.633	0.561	0.531
Informal – moral suasion	Male sample	191	0	0.539	0.482	0.466
	Female sample	196	100	0.684	0.582	0.577
Formal – moral duty	Mixed	175	5.7	0.611	0.543	0.509
Informal – moral duty	Male sample	194	0	0.711	0.629	0.603
	Female sample	169	100	0.740	0.669	0.639
Formal – social norm	Mixed	185	14.6	0.605	0.530	0.514
Informal – social norm	Male sample	176	0	0.767	0.648	0.625
	Female sample	193	100	0.788	0.632	0.622

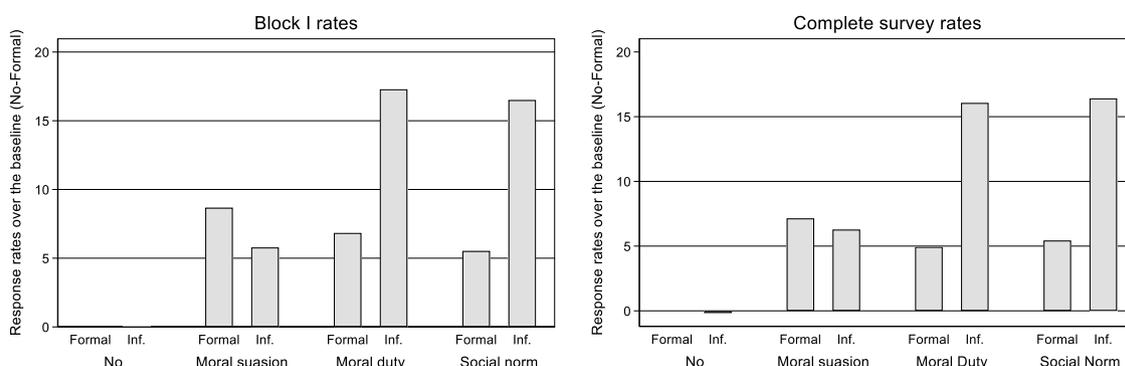


Fig. 1. Differences in response rates over baseline by treatment.

Table 3
Treatment effects for the formal treatments over the baseline.

	Formal Treatments			
	Baseline	Moral suasion	Moral duty	Social norm
Block I	0.474	0.561**	0.543*	0.530
Completed	0.459	0.531*	0.509	0.514
% female	14.3	14.3	5.7	14.6

*, **, *** = 10%, 5%, 1% significance with respect to baseline treatment (i.e. formal-no norm).

^, ^^, ^^ = 10%, 5%, 1% significance with respect to baseline treatment after applying Bonferroni correction.

2.3. Hypotheses

Based on the literature reviewed above, two hypotheses were proposed regarding the effects of the two main treatment dimensions' variables: writing style and norm information. Both kinds of variables were expected to have a positive impact on survey participation as follows:

Hypothesis 1: Information provided about norms will promote participation in the survey.

Hypothesis 2: Informal messages including the recipients' own name will promote participation in the survey.

An interesting issue was the gender dimension's impact on the nudges' efficacy. The economics literature includes many studies of how gender affects preferences, such as social preferences (Andreoni & Vesterlund, 2001), preferences for competition (Niederle & Vesterlund, 2007) or risk preferences (Charness & Gneezy, 2012). However, whether responses to nudging differ by gender is a topic that has not yet received much attention. One exception is prior research on how informational nudges affect retirement saving decisions differently for male versus female workers (Clark et al., 2014). Although the present study was not originally designed to analyse any potential differential effects of nudges by gender, the need to compose for and send emails to male and female

groups facilitated this type of analysis. Hypothesis 3 captured the most conservative view of no gender-related differences:

Hypothesis 3: Women and men will react equally to norm nudges used to promote the survey.

3. Results

3.1. Descriptive results

Table 2 displays the response rates across the 12 samples for three possible outcomes. These are (1) faculty who clicked on the link but did not proceed any further or Click outcome, (2) respondents who completed the survey's first block or Block I outcome and, (3) staff members who answered all five blocks of questions or Completed outcome. The decay in response rates across the sequence Click-Block I-Completed is quite uniform across treatments. The average ratio of Block I to Click response rates is 0.8639 (standard deviation [SD] 0.0345), while the average ratio is 0.9638 (SD 0.0152) when moving from Block I to Completed. Analyses were conducted with reference to two outcomes: the response rate for Block I, which is the outcome the QAU had used in previous years to evaluate the online survey's success, and the rate for completed surveys.⁵

From Table 2, we observe differences in response rates by norms and writing styles. To get a sense of the data, in Fig. 1 we pool the two gendered versions of each of the informal messages and display the difference in answer rates over the baseline (Formal - No norm).

In the baseline treatment, the response rates are 0.474 (Block I) and

⁵ See Appendix B for Tables B3 and B4, which show the results for analyses of the treatment and gender effects for the remaining outcome (i.e. clicked surveys). The treatment effects are quite robust for the outcome in question.

Table 4

Treatment effects for the informal treatments (adjusted for gender composition) over the baseline.

	Informal Treatment			
	No	Moral suasion	Moral duty	Social norm
Block I	0.433	0.496	0.635*** ^^^	0.646*** ^^^
Completed	0.414	0.482	0.608*** ^^^	0.625*** ^^^
% female	14.3	14.3	14.3	14.3

*, **, *** = 10%, 5%, 1% significance with respect to baseline treatment (i.e. formal-no norm).

^, ^^, ^^^ = 10%, 5%, 1% significance with respect to baseline treatment after applying Bonferroni correction.

0.459 (Completed).⁶ The treatments that yield a larger increase in response rates with regard to this treatment are moral duty and social norm – both using an informal writing style.

3.2. Statistical analysis

We start assessing whether there are significant differences among the treatments. The comparison between the formal style coupled with a norm over the baseline treatment can be done using standard proportion tests as the fraction of female respondents is around 14%, similar to the female ratio in the baseline, so the gender variable is not a confounding factor.⁷ Table 3 displays the results.

We observe that moral suasion presents a significant increase in answer rates for Block I over the baseline treatment and a marginally significant increase for the Completed outcome. Moral duty also produces an only marginally significant increase in rates for Block I. However, a more conservative approach was next applied in which the Bonferroni correction for multiple hypothesis testing was used.⁸ Among all the effects quantified above, no one remains statistically significant.

Result 1: We find no significant differences among the formal treatments.

A different picture emerges when an informal writing style is used. As the analysis reported in Tables B1 and B2 (in Appendix B) reveals, with informal writing two of the three norms provide significantly higher response rates than that of the informal-no norm treatment, namely, informal-moral duty and informal-social norm.⁹ These effects remain significant even after using the Bonferroni correction for multiple hypotheses testing, which constitutes our second result:

Result 2: The provision of information about some norms – specifically moral duty and social norm – has a positive effect on faculty satisfaction

⁶ A test of proportions was run to compare the Block I response rate of the baseline treatment (0.474) to the preceding academic year's rate (0.517), revealing that the difference is not statistically significant ($z = 1.037$; two-sided p -value = 0.2997).

⁷ This holds for all treatments except for the formal-moral duty, for which according to the data provided by the QAU, the female ratio is 5.7%.

⁸ This research's design included eight treatments, and multiple hypotheses were tested, which increased the probability of obtaining false positive results. By applying the Bonferroni correction for multiple hypotheses testing, the critical p -values could be divided by eight. The corrected critical p -values are as follows: $0.1/8 = 0.0125$ for a significance level of 10% (i.e. ^ or marginally significant); $0.05/8 = 0.00625$ for a significance level of 5% (i.e. ^^ or significant); and $0.01/8 = 0.00125$ for a significance level of 1% (i.e. ^^^ or highly significant).

⁹ For the Block I outcome (see Table B1 in Appendix B), the moral suasion treatment produces a difference of 0.057, and the one-sided p -value is 0.0540. The moral duty treatment's difference is 0.172, and the one-sided p -value is 0.0000. The social norm treatment generates a difference of 0.165, and the one-sided p -value is 0.0000. For the Completed outcome (see Table B2 in Appendix B), the moral suasion treatment is associated with a difference of 0.064, and the one-sided p -value is 0.0353. The moral duty treatment's difference is 0.162, and the one-sided p -value is 0.0000. The social norm treatment produces a difference of 0.166, and the one-sided p -value is 0.0000.

survey participation when coupled with informal writing.

Results 1 and 2 provides partial support for Hypothesis 1. The most successful treatments (i.e. moral duty and social norm with an informal writing style) have the same effect on response rates for both the Block I and Completed outcomes, as shown in Tables B1 and B2 (Appendix B).¹⁰ The following result is thus as follows:

Result 3: Moral duty and social norm coupled with an informal writing style offer statistically equivalent results as these nudges and style together produce the strongest treatment effects.

For the analysis of the effect of informal writing on response rates, Hypothesis 2, we cannot make direct comparisons because the female gender composition of the formal samples is around 14% and we do not have any informal samples whose female composition is 14%. Rather, we do have two gendered informal samples for each norm, so we can apply a bootstrapping technique to generate informal samples with a female gender composition of 14%.¹¹ The p -values were computed by applying a bootstrap procedure with 10,000 iterations, in which the informal treatments' generated samples had exactly the same gender composition as the formal treatment sample (i.e. 14% females). Table 4 report the treatment effects and their significance levels over the baseline treatment.¹²

The informal treatments implying moral duty and social norm yield significantly higher response rates than the baseline does, even after the Bonferroni correction is made, both at the 1% level. This constitutes the next result:

Result 4: The behavioural intervention applied – specifically informal-moral duty and informal-social norm treatments – has a positive impact on faculty satisfaction survey participation, significantly increasing response rates compared to the baseline.

Table 2 shows interesting interaction effects between the two dimensions (i.e. norms and writing style). To assess the informal writing style's impact over the formal one, we will also use the informal treatments' generated samples that had the same gender composition as the formal ones. Our analysis, reported in Tables B1 and B2 (in Appendix B), reveals that this effect is positive and significant only for the moral duty and social norm treatments.¹³ When a more conservative approach is used based on the Bonferroni correction for multiple hypotheses testing (cf. footnote 8), the informal writing style's effect only remains

¹⁰ For the Block I outcome (see Table B1 in Appendix B), the difference is 0.008, and the one-sided p -value is 0.4126. For the Completed outcome (see Table B2 in Appendix B), the difference is 0.003, and the one-sided p -value is 0.4614.

¹¹ Bootstrap techniques are used in the analysis of experimental data. See for example Section 2.8 of Moffatt (2021), which surveys a range of econometric techniques that are currently being used by experimental economists. Van Doorslaer and Koolman (2004) use a similar bootstrap (resampling) approach to ours, which is used to statistically test for cross-country differences in the degree of income-related inequalities in self-assessed health. In Doorslaer and Koolman's setup, given the complexity of their survey design, they need to assess sampling variability in terms of age and gender. Other empirical studies using a bootstrap methodology to test for differences are, for instance, Boisso et al. (1994) and Hellerstein and Neumark (2008), which study occupational and workplace segregation. In this case Hellerstein and Neumark use the bootstrap methodology to compare across samples that have some overlap in terms of race and language.

¹² The p -values for the tests are given in Tables B1 and B2 in Appendix B, which also report the p -values for all the possible pairwise comparisons for the Block I and Completed outcomes, respectively. Table B3 further lists the p -values for the Completed outcome.

¹³ For the Block I outcome (see Table B1 in Appendix B), the moral duty treatment produces a difference of 0.088, and the one-sided p -value is 0.0409. For the social norm treatment, the difference is 0.116 and the one-sided p -value is 0.0099. For the Completed outcome (see Table B2 in Appendix B), the moral duty treatment's difference is 0.097, and the one-sided p -value is 0.0328. The social norm treatment generates a difference of 0.111, and the one-sided p -value is 0.0138.

Table 5
Gender's effect on response rates for informal treatments.

a. Block I rates					
Treatment	Frequency females	Frequency males	Difference	P-value One-sided	Two-sided
Informal-no norm	0.535	0.416	0.120	0.0084	0.0168
Informal-moral suasion	0.582	0.482	0.100	0.0244	0.0488
Informal-moral duty	0.669	0.629	0.040	0.2144	0.4288
Informal-social norm	0.632	0.648	-0.016	0.3776	0.7552
b. Completed questionnaires rates					
Treatment	Frequency females	Frequency males	Difference	P-value One-sided	Two-sided
Informal-no norm	0.520	0.396	0.124	0.0063	0.0127
Informal-moral suasion	0.577	0.466	0.111	0.0147	0.0295
Informal-moral duty	0.639	0.603	0.036	0.2407	0.4814
Informal-social norm	0.622	0.625	-0.003	0.4744	0.9488

significant for the social norm treatment, which comprises the fifth result:

Result 5: Informal writing alone does not promote participation in faculty satisfaction online surveys, but this style has a significant impact when accompanied by an appeal to a social norm, which provides partial support for Hypothesis 2.

3.3. Robustness analysis

To test the results' robustness, logit econometric analysis was also performed. The probability of filling in the questionnaire for both the Block I and Completed outcomes was regressed as a function of dummy variables representing treatment effects (i.e. moral suasion, moral duty and social norm). This analysis further included a dummy variable for the writing style variable (i.e. informal), which also interacted with the treatment variables in the regression.

Because the questionnaire asked respondents to complete an item about their position at the university, this variable could be controlled for in the regressions.¹⁴ An explanatory variable was thus integrated as a dummy variable (i.e. permanent) indicating whether the respondents had tenure at the university. In addition, the analysis incorporated this variable's interactions with the treatment and writing style variables.

The results are shown in Appendix C. Tables C1 lists the estimated values, while Table C2 presents the treatment variables' (i.e. norms and writing style) marginal effects. This table more specifically highlights the analogous results for the two outcomes considered (i.e. Block I and Completed). Table C3, in turn, provides details on tenure's marginal effects. The latter table's contents are discussed further in the next Section.

The moral suasion, moral duty and social norm treatments' significant marginal effects are detectable only when an informal writing style was used. This style's marginal effect is significant only in association

with the moral duty and social norm treatments. With the more conservative approach using the Bonferroni correction for multiple hypotheses testing, the only marginal effects that remain significant are the moral duty and social norm treatments (i.e. when an informal writing style is used) and the informal writing style with social norm treatment. Therefore, the econometric analyses confirmed that the results are empirically robust after controlling for the respondents' position at the university (i.e. tenure effects).

3.4. Gender, tenure and reminder effects

For each treatment involving an informal writing style, the QAU generated two subsamples: one of all-male faculty and another of all-female academic staff. These samples facilitated an examination of gender's effect on response to norm nudges in the informal treatments.¹⁵

Table 5 displays the results of standard probability tests. When either moral suasion or no norm is used, female faculty have significantly higher response rates than men do. Female respondents are thus more prone to participate in the survey in the absence of norms (i.e. intrinsic motivation) or in the presence of a moral suasion message.¹⁶ However, this gender effect disappears in the two most successful treatments – moral duty and social norm – which are equally appealing to both males and females. This result can be expressed as follows:

Result 6: Informal writing without any norm nudge and with moral suasion induces women to participate in faculty satisfaction surveys slightly more often than men do, but the gender effect vanishes with moral duty and social norm nudging, thereby providing partial support for Hypothesis 3.

Appendices D and E contain the results of further analyses of tenure and response time's potential effects, respectively. In the 2016–2017 academic year, 60% of the faculty members held a tenured position. The information used to differentiate between tenured and untenured faculty in this intervention was self-reported by respondents in the survey's Block I. At the time the survey was conducted, 58% of the faculty were permanent. Table D1 in Appendix D specifically shows the percentage of permanent faculty at the University of Malaga by treatment. In most treatments, this figure was close to the average, except for one treatment in which the tenured faculty were under-represented (i.e. the informal-social norm).

Table D2, in turn, lists the response rates for both tenure categories. For all the treatments and outcomes considered (i.e. Block I and Completed), the permanent faculty's response rates are always higher than those of the non-tenured faculty. The econometric analysis covered in Appendix C assessed how significant these tenure effects are. As Table C3 shows, having a tenured position has a significant positive impact on all eight treatments. When the Bonferroni correction is applied, all these effects remain significant, with the exception of the formal-social norm treatment.

Since the non-permanent faculty have lower response rates than the tenured staff members do, the under-representation of the latter faculty identified in the informal-social norm treatment should counteract this treatment's positive effect. However, the data do not follow the expected pattern because the treatment achieves the highest response rate in the field experiment (i.e. in tandem with the informal-moral duty treatment).

Regarding response time, the survey software only timestamped the moment the questionnaire was completed, so the analysis focused on the Completed outcome. Fig. E1 in Appendix E plots the response rates in blocks of three hours after the survey is launched (i.e. see the top panel) and the histogram of rates after the initial launch and each reminder, also in blocks of three hours (i.e. see the bottom panel). The results

¹⁴ The questionnaire had no item asking respondents to report their gender. Since the formal treatments involved a mixed gender sample, this variable could not be controlled for in the regressions.

¹⁵ For the formal treatments, this analysis cannot be performed.

¹⁶ Notably, when the Bonferroni correction is used, the critical *p*-values become smaller, and the differences are only marginally significant.

suggest that the faculty members either react to the invitation and/or reminder to fill in the questionnaire as soon as they receive it, or they do not react at all.

Overall, reminders are decreasingly useful in terms of increasing participation, as can be seen in Fig. E2 of Appendix E, which plots the cumulative frequency of the received surveys as a function of the number of hours elapsed since the initial invitation is sent. A visual inspection of this figure reveals shifts of decreasing size in the frequency with each reminder. These shifts present the same pattern across treatments (see Fig. E3) based on the cumulative distribution across norms. Fig. E4 further shows the distribution across writing styles (see Appendix E).

4. Discussion and conclusion

The motivation for conducting the 2017 intervention at the University of Malaga was the need to investigate how using insights from the behavioural sciences can affect participation rates in this institution's quality assurance system. Overall participation in the 2016–2017 academic year (i.e. 55%) was significantly higher than in the pre-intervention year (test of proportion: $z = 2.951$; one-sided p -value = 0.0016). Thus, the intervention was a success because it increased participation in the faculty satisfaction survey by 4 percentage points.

The results of this field experiment offer two insights. First, a moral duty nudge and a peer information nudge providing recipients with details about other people's behaviour (i.e. social norm treatment) have the same greatest impact on response rates – as long as an informal writing style is used. This finding can be seen as good news because, when nudges based on social norms are used, quite frequently only a minority of the target population will engage in the behaviour being promoted. This tendency precludes the use of messages such as 'the majority of' or 'most of'. In addition, appeal letters should not place 'a minority of' after the norm since this message typically reinforces the opposite of the desired behaviour (Bicchieri & Dimant, 2019). The present intervention confirmed that an alternative approach is to avoid relying on information about peers' behaviour and instead appealing to recipients by making participation a moral duty, which can have a similar positive impact.

Second, this field experiment revealed that nudging can have gender-specific effects. This result is quite interesting in isolation, but it could complicate the use of nudges in generalised settings and academic institutions, in particular, as segmentation by gender may need to be applied when using nudges. This finding's positive side is that the two most successful treatments are not associated with gender-specific effects. That is, regardless of a target population's gender composition, social norms and moral duty are equally appealing to all, which further simplifies the formulation of messages. However, any extrapolation of these results should take into account that some data contamination might have occurred since respondents from the same university school could have influenced each other's decisions, which could not be controlled for in the analyses (cf. footnote 3).

The collaboration with the University of Malaga's QAU ended just after the final report was submitted in September 2017. This report contained clear recommendations for how to use the most successful treatments in the following academic year. The QAU followed that advice and implemented the informal-moral duty approach in their message sent in June 2018. The response rates further increased, reaching 59.1%, which was significantly higher than the previous year's average of 55.6% ($z = 2.4652$; $p = 0.0068$), although this percentage was somewhat less than the 64.7% that the informal-moral duty treatment had achieved in 2017 ($z = 2.0287$; $p = 0.0212$).

In June 2019, the QAU did not use the informal-moral duty

message, returning instead to the formal-no norm pre-intervention email, but with two modifications. First, because the questionnaire was rather long, the unit provided faculty members with four links – one for each part of the survey – rather than the customary single link to the entire questionnaire. Second, the links were sent out without a token, so faculty members who had not yet filled out the survey did not receive personalised reminders. The result was a decline in the participation rate down to 25% – the lowest response rate in the previous six years. Although an analysis cannot be carried out of these two changes' role in this drop in the faculty's engagement, placing blame on the modifications would not be unreasonable given the response rate's sharp decrease from 60% to 25%.

Appendix A. Email messages

Spanish translation

Asunto: Invitación para participar en la encuesta
[Estimado/a Sr./Sra.] [Hola #Antonio] [Hola #Ana]:
[Ha] [Has] [Has] sido [invitado] [invitado] [invitada] a participar en la siguiente encuesta: «Cuestionario de Satisfacción de Profesores del Sistema de Garantía de la Calidad (2016–2017)»
[Moral suasion treatment]: Necesitamos **[su] [tu] [tu]** colaboración para que la UMA mejore. Cuantos más completemos el cuestionario, mayores oportunidades de mejora.
[Moral duty treatment]: Uno de **[sus] [tus] [tus]** deberes como PDI es colaborar para que la UMA mejore.
[Social norm treatment]: La mayoría de los profesores de la UMA rellenan esta encuesta.
 Para hacerlo, por favor **[pulse] [pulsa] [pulsa]** en el siguiente enlace:
[survey link].
 Muchas gracias por **[su] [tu] [tu]** interés y colaboración.
 Atentamente,
[null] [Eva] [Eva]
 Servicio de Calidad, Planificación Estratégica y Responsabilidad Social

English translation

Subject: Invitation to participate in faculty satisfaction survey
[Dear Sir/Madam] [Hi #Antonio] [Hi #Ana],
 You have been invited to participate in the Satisfaction Survey of the Quality Assurance System (2016–2017).
[Moral suasion treatment]: We need your help to improve the University of Malaga. If more of us complete the questionnaire, we will have more opportunities to make improvements.
[Moral duty treatment]: One of your duties as a faculty member is to help the University of Malaga improve.
[Social norm treatment]: A majority of the academic staff at the University of Malaga completes this survey.
 To do so, please click on the following link: **[survey link].**
 Thank you very much for your time and participation.
 Yours sincerely,
[null] [Eve] [Eve]
 The QAU
Instructions for how to compose the message
 On the lines with three options in squared brackets, the first one refers to the formal treatment, the second to the informal male version and the third one to the informal female version: **[formal] [informal male] [informal female].**

Appendix B. Results for analyses of treatment effects for three outcomes – Block I, Completed and Click – and gender effects for Click outcome (probability tests)

Table B1

Treatment effects on frequency of questionnaires with first block completed (Block I).

Treatment T _B	Treatment T _A	Freq. T _B	Freq. T _A	Diff.	P-value One- sided	Two- sided
No informal	No formal	0.433	0.474	-0.042	0.2209 [#]	0.4418 [#]
Moral	Moral	0.496	0.561	-0.065	0.1022 [#]	0.2044 [#]
suasion informal	suasion formal					
Moral duty informal	Moral duty formal	0.631	0.543	0.088	0.0409 [#]	0.0818 [#]
Social norm informal	Social norm formal	0.645	0.530	0.116	0.0099 [#]	0.0198 [#]
Moral	No formal	0.561	0.474	0.087	0.0429	0.0858
suasion formal						
Moral duty formal	No formal	0.543	0.474	0.068	0.0943	0.1886
Social norm formal	No formal	0.530	0.474	0.055	0.1406	0.2812
Moral duty formal	Moral suasion formal	0.543	0.561	-0.018	0.3612	0.7224
Social norm formal	Moral suasion formal	0.530	0.561	-0.031	0.2686	0.5372
Social norm formal	Moral duty formal	0.530	0.543	-0.013	0.4014	0.8028
Moral	No formal	0.496	0.474	0.021	0.3201 [#]	0.6402 [#]
suasion informal						
Moral duty informal	No formal	0.635	0.474	0.160	0.0006 [#]	0.0012 [#]
Social norm informal	No formal	0.645	0.474	0.171	0.0002 [#]	0.0004 [#]
Moral	No informal	0.532	0.475	0.057	0.0540 ^{&}	0.1080 ^{&}
suasion informal						
Moral duty informal	No informal	0.647	0.475	0.172	0.0000 ^{&}	0.0000 ^{&}
Social norm informal	No informal	0.640	0.475	0.165	0.0000 ^{&}	0.0000 ^{&}
Moral duty informal	Moral suasion informal	0.647	0.532	0.115	0.0007 ^{&}	0.0014 ^{&}
Social norm informal	Moral suasion informal	0.640	0.532	0.107	0.0014 ^{&}	0.0028 ^{&}
Social norm informal	Moral duty informal	0.640	0.647	-0.008	0.4126 ^{&}	0.8252 ^{&}

[#] Bootstrap procedure with 10,000 iterations, in which generated informal treatment (T_B) samples have exactly the same gender composition as formal treatment (T_A) samples, used to compute test's *p*-values.

[&] For the comparisons between informal treatments, we have pooled the all-female and all-male samples of each norm, yielding a gender-balanced sample.

Table B2

Treatment effects on frequency of fully completed questionnaires (Completed).

Treatment T _B	Treatment T _A	Freq. T _B	Freq. T _A	Diff.	P-value One- sided	Two- sided
No informal	No formal	0.414	0.459	-0.045	0.1926 [#]	0.3852 [#]
Moral	Moral	0.482	0.531	-0.049	0.1738 [#]	0.3476 [#]
suasion informal	suasion formal					
Moral duty informal	Moral duty formal	0.605	0.509	0.097	0.0328 [#]	0.0656 [#]

Table B2 (continued)

Treatment T _B	Treatment T _A	Freq. T _B	Freq. T _A	Diff.	P-value One- sided	Two- sided
Social norm informal	Social norm formal	0.625	0.514	0.111	0.0138 [#]	0.0276 [#]
Moral	No formal	0.531	0.459	0.071	0.0786	0.1572
suasion formal						
Moral duty formal	No formal	0.509	0.459	0.049	0.1710	0.3420
Social norm formal	No formal	0.514	0.459	0.054	0.1445	0.2890
Moral duty formal	Moral suasion formal	0.509	0.531	-0.022	0.3357	0.6714
Social norm formal	Moral suasion formal	0.514	0.531	-0.017	0.3692	0.7384
Social norm formal	Moral duty formal	0.514	0.509	0.005	0.4627	0.9254
Moral	No formal	0.482	0.459	0.023	0.3140 [#]	0.6280 [#]
suasion informal						
Moral duty informal	No formal	0.608	0.459	0.149	0.0009 [#]	0.0018 [#]
Social norm informal	No formal	0.625	0.459	0.165	0.0005 [#]	0.001 [#]
Moral	No informal	0.522	0.457	0.064	0.0353 ^{&}	0.0706 ^{&}
suasion informal						
Moral duty informal	No informal	0.620	0.457	0.162	0.0000 ^{&}	0.0000 ^{&}
Social norm informal	No informal	0.623	0.457	0.166	0.0000 ^{&}	0.0000 ^{&}
Moral duty informal	Moral suasion informal	0.620	0.522	0.098	0.0034 ^{&}	0.0068 ^{&}
Social norm informal	Moral suasion informal	0.623	0.522	0.101	0.0024 ^{&}	0.0048 ^{&}
Social norm informal	Moral duty informal	0.623	0.620	0.003	0.4614 ^{&}	0.9228 ^{&}

[#] Bootstrap procedure with 10,000 iterations, in which generated informal treatment (T_B) samples have exactly the same gender composition as formal treatment (T_A) samples, used to compute test's *p*-values.

[&] For the comparisons between informal treatments, we have pooled the all-female and all-male samples of each norm, yielding a gender-balanced sample.

Table B3

Treatment effects on frequency of clicked questionnaires (Click).

Treatment T _B	Treatment T _A	Freq. T _B	Freq. T _A	Diff.	P-value One- sided	Two- sided
No informal	No formal	0.527	0.566	-0.039	0.2262 [#]	0.4524 [#]
Moral	Moral	0.560	0.633	-0.073	0.0720 [#]	0.1440 [#]
suasion informal	suasion formal					
Moral duty informal	Moral duty formal	0.713	0.611	0.102	0.0194 [#]	0.0388 [#]
Social norm informal	Social norm formal	0.770	0.605	0.165	0.0004 [#]	0.0008 [#]
Moral	No formal	0.633	0.566	0.066	0.0901	0.1802
suasion formal						
Moral duty formal	No formal	0.611	0.566	0.045	0.1892	0.3784
Social norm formal	No formal	0.605	0.566	0.039	0.2195	0.439
Moral duty formal	Moral suasion formal	0.611	0.633	-0.021	0.3369	0.6738
Social norm formal	Moral suasion formal	0.605	0.633	-0.027	0.2920	0.5840
Social norm formal	Moral duty formal	0.605	0.611	-0.006	0.4534	0.9068

Table B3 (continued)

Treatment T _B	Treatment T _A	Freq. T _B	Freq. T _A	Diff.	P-value One- sided	Two- sided
Moral suasion informal	No formal	0.560	0.566	-0.006	0.4592#	0.9184#
Moral duty informal	No formal	0.715	0.566	0.149	0.0005#	0.0010#
Social norm informal	No formal	0.770	0.566	0.204	0.0000#	0.0000#
Moral suasion informal	No informal	0.612	0.558	0.055	0.0591&	0.1182&
Moral duty informal	No informal	0.725	0.558	0.167	0.0000&	0.0000&
Social norm informal	No informal	0.778	0.558	0.220	0.0000&	0.0000&
Moral duty informal	Moral suasion informal	0.725	0.612	0.112	0.0006&	0.0012&
Social norm informal	Moral suasion informal	0.778	0.612	0.165	0.0000&	0.0000&
Social norm informal	Moral duty informal	0.778	0.725	0.053	0.0478&	0.0956&

Bootstrap procedure with 10,000 iterations, in which generated informal treatment (T_B) samples have exactly the same gender composition as formal treatment (T_A) samples, used to compute test's p-values.

& For the comparisons between informal treatments, we have pooled the all-female and all-male samples of each norm, yielding a gender-balanced sample.

Table B4

Gender effects on frequency of clicked questionnaires (Click).

Treatment	Frequency females	Frequency males	Difference	P-value One- sided	Two- sided
No norm informal	0.601	0.515	0.086	0.0414	0.0828
Moral suasion informal	0.684	0.539	0.144	0.0018	0.0036
Moral duty informal	0.740	0.711	0.028	0.2735	0.5470
Social norm informal	0.788	0.767	0.021	0.3179	0.6358

Appendix C. Econometric analysis of treatment effects

Table C1

Logit regressions.

Independent variables	Block I	Completed
Informal	-0.086 (0.2170)	-0.098 (0.2173)
Permanent	0.735*** (0.2170)	0.693*** (0.2173)
Moral_suasion	0.357 (0.2585)	0.248 (0.2594)
Moral_duty	0.376 (0.2589)	0.291 (0.2593)
Social_norm	0.434* (0.2594)	0.408 (0.2597)
Moral_suasion × Informal	-0.086 (0.2529)	0.007 (0.2526)
Moral_duty × Informal	0.404 (0.2621)	0.434* (0.2607)
Social_norm × Informal	0.594** (0.2606)	0.596** (0.2601)

Table C1 (continued)

Independent variables	Block I	Completed
Moral_suasion × Permanent	0.029 (0.2490)	0.099 (0.2498)
Moral_duty × Permanent	-0.008 (0.2554)	-0.008 (0.2546)
Social_norm × Permanent	-0.283 (0.2546)	-0.248 (0.2541)
Informal × Permanent	0.088 (0.1876)	0.094 (0.1873)
Constant	-0.572*** (0.2024)	-0.608*** (0.2031)
Observations	2,271	2,271

Note. Permanent = tenured.

Table C2

Marginal effects of different treatments and writing style.

Independent variables	Block I	Completed
Moral_suasion	with formal (0.0495)	0.074 (0.0495)
	with informal (0.0350)	0.075** (0.0349)
Moral_duty	with formal (0.0511)	0.069 (0.0513)
	with informal (0.0350)	0.172*** (0.0352)
Social_norm	with formal (0.0506)	0.064 (0.051)
	with informal (0.0354)	0.205*** (0.0356)
Informal	with no norm (0.0427)	-0.010 (0.0427)
	with moral_suasion (0.0429)	-0.009 (0.0429)
	with moral_duty (0.0447)	0.092** (0.0452)
	with social_norm (0.0443)	0.131*** (0.0446)

* ** *** = 10%, 5%, 1% significance.

^, ^, ^ = 10%, 5%, 1% significance using Bonferroni correction.

Table C3

Marginal effects of tenure (Permanent).

Independent variables	Block I	Completed
With no norm	and formal (0.0514)	0.169*** (0.0513)
	and informal (0.0442)	0.190*** (0.0439)
With moral_suasion	and formal (0.0515)	0.195*** (0.0514)
	and informal (0.0442)	0.218*** (0.0440)
With moral_duty	and formal (0.0528)	0.169*** (0.0530)
	and informal (0.0447)	0.185*** (0.0453)
With social_norm	and formal (0.0535)	0.111** (0.0533)
	and informal (0.0425)	0.124*** (0.0430)

* ** *** = 10%, 5%, 1% significance.

^, ^, ^ = 10%, 5%, 1% significance using Bonferroni correction.

Appendix D. Tenure effects

Table D1

Percentage of permanent faculty by treatment.

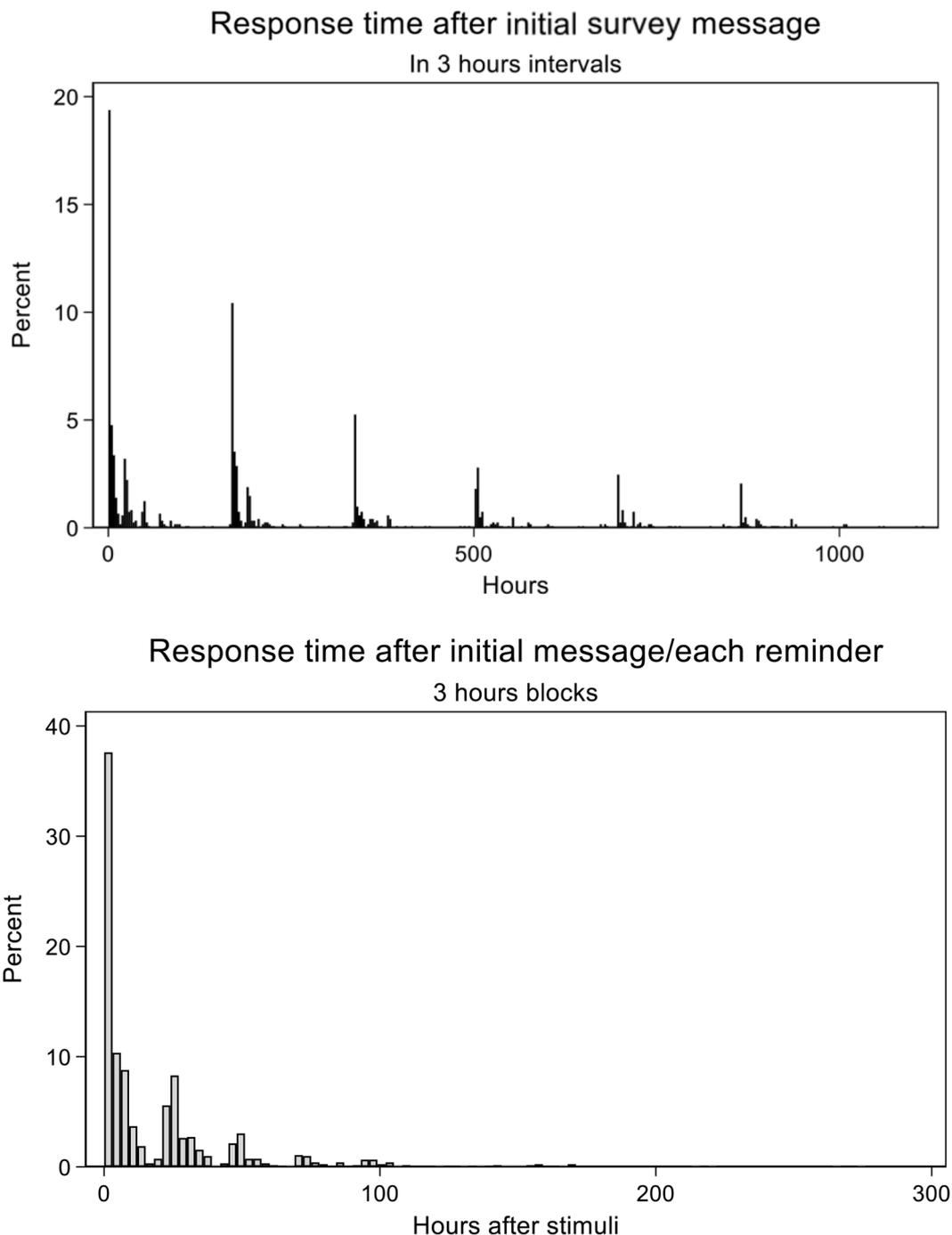
	No norm	Moral suasion	Moral duty	Social norm
Formal	63.18	61.19	51.10	57.14
Informal	67.00	60.95	61.98	39.72

Note. Permanent = tenured.

Table D2
Response rates for Block I and Completed by treatment and tenure category.

Tenure category	Outcome	Writing style	Norm No norm	Moral suasion	Moral duty	Social norm
Permanent	Block I	Formal	0.524	0.658	0.630	0.576
		Informal	0.549	0.602	0.721	0.714
	Completed	Formal	0.500	0.626	0.584	0.576
		Informal	0.530	0.598	0.695	0.687
Non- permanent	Block I	Formal	0.389	0.408	0.459	0.468
		Informal	0.326	0.424	0.526	0.590
	Completed	Formal	0.389	0.382	0.435	0.430
		Informal	0.311	0.404	0.496	0.581

Appendix E. Response time –effects of reminders on response rates



10
Fig. E1. Response times.

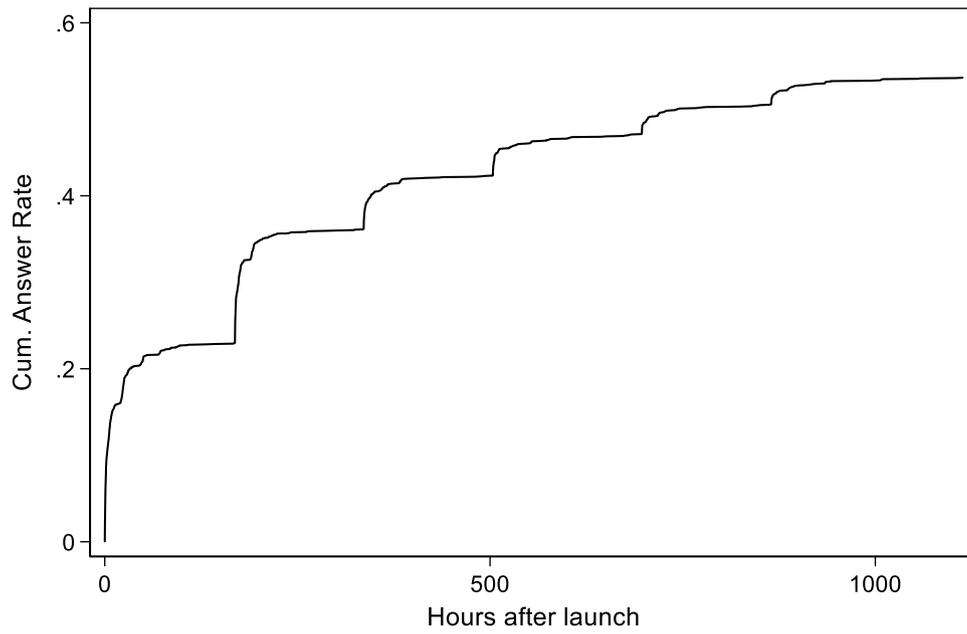


Fig. E2. Cumulative answer rate (completed surveys) over time.

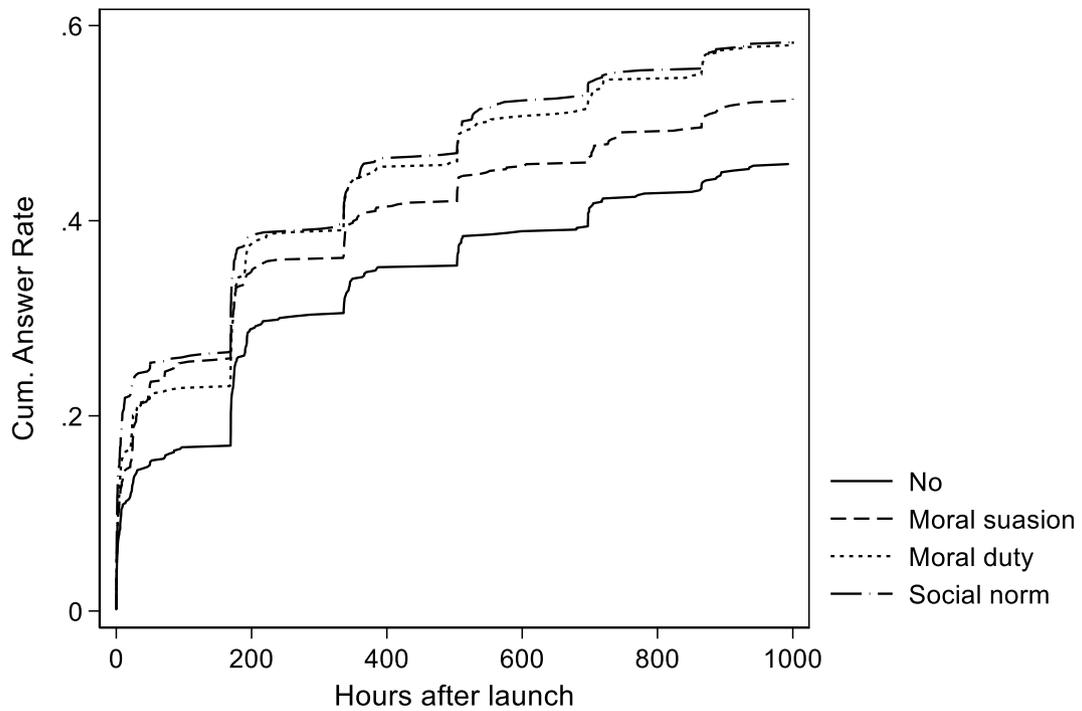


Fig. E3. Cumulative answer rate (completed surveys) over time by norm.

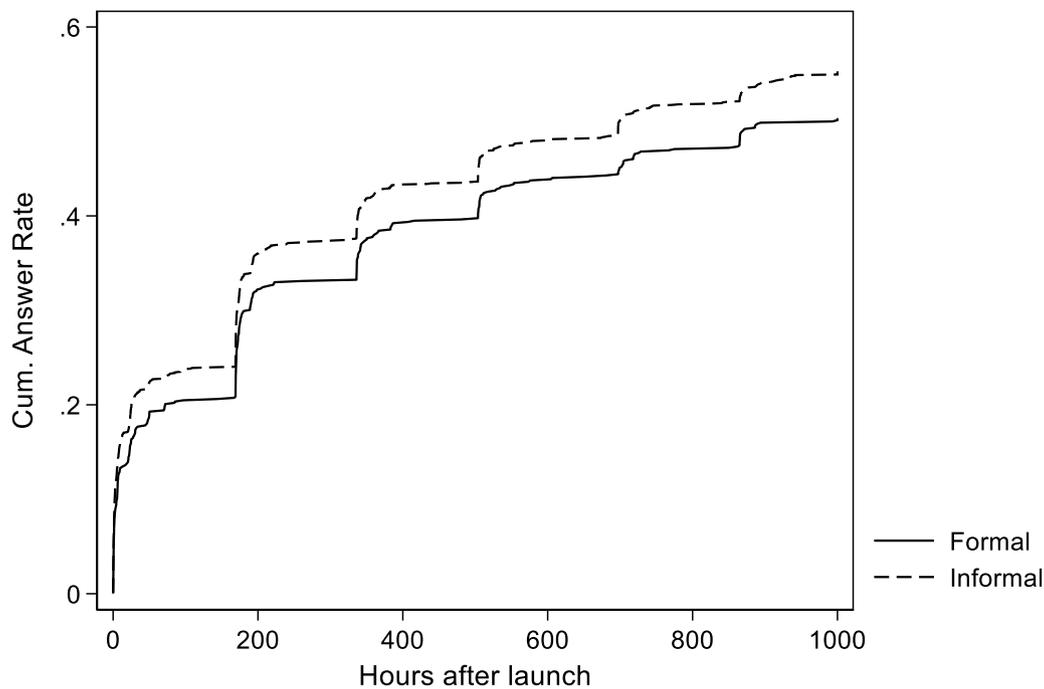


Fig. E4. Cumulative answer rate (completed surveys) over time by writing style.

References

- Allcott, H. (2011). Special issue: The role of firms in tax systems. Social norms and energy conservation *Journal of Public Economics*, 95, 1082–1095. <https://doi.org/10.1016/j.jpubeco.2011.03.003>.
- Alm, J., & Torgler, B. (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics*, 101(4), 635–651.
- Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Q J Econ*, 116, 293–312. <https://doi.org/10.1162/003355301556419>
- Baumruk, R. (2006). Why managers are crucial to increasing engagement: Identifying steps managers can take to engage their workforce. *Strategic HR Review*, 5, 24–27. <https://doi.org/10.1108/14754390680000863>
- Bedarkar, M., & Pandita, D. (2014). A study on the drivers of employee engagement impacting employee performance. *Procedia-Social and Behavioral Sciences*, 133, 106–115.
- Behavioural Insights Team, 2012. Fraud, error and debt: behavioural insights team paper [WWW Document]. GOV.UK. URL <https://www.gov.uk/government/publications/fraud-error-and-debt-behavioural-insights-team-paper> (accessed 11.18.20).
- Behavioural Insights Team, 2013. Applying behavioural insights to charitable giving [WWW Document]. GOV.UK. URL <https://www.gov.uk/government/publications/applying-behavioural-insights-to-charitable-giving> (accessed 11.18.20).
- Behavioural Insights Team, 2014. EAST: Four Simple Ways to Apply Behavioural Insights [WWW Document]. URL <https://www.bi.team/publications/east-four-simple-ways-to-apply-behavioural-insights/> (accessed 11.18.20).
- Bicchieri, C., & Dimant, E. (2019). Nudging with care: the risks and benefits of social information. *Public Choice*. <https://doi.org/10.1007/s11127-019-00684-6>
- Blumenthal, M., Christian, C., Slemrod, J., & Smith, M. G. (2001). Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota. *National Tax Journal*, 54, 125–138.
- Boisso, D., Hayes, K., & Hirschberg, J. (1994). Occupational segregation in the multidimensional case: Decomposition and tests of significance. *Journal of Econometrics*, 61, 161–171.
- Bott, K. M., Cappelen, A. W., Sørensen, E.Ø., & Tungodden, B. (2019). You've got mail: A randomized field experiment on tax evasion. *Management Science*, 66, 2801–2819. <https://doi.org/10.1287/mnsc.2019.3390>
- Capraro, V., & Vanzo, A. (2019). The power of moral words: Loaded language generates framing effects in the extreme dictator game. *Judgment and Decision Making*, 14, 309–317.
- Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83, 50–58. <https://doi.org/10.1016/j.jebo.2011.06.007>
- Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica*, 77, 909–931. <https://doi.org/10.3982/ECTA7416>
- Clark, R. L., Maki, J. A., & Morrill, M. S. (2014). Can simple informational nudges increase employee participation in a 401(k) Plan? *Southern Economic Journal*, 80, 677–701. <https://doi.org/10.4284/0038-4038-2012.199>
- Dal Bó, E., & Dal Bó, P. (2014). Do the right thing? The effects of moral suasion on cooperation. *Journal of Public Economics*, 117, 28–38. <https://doi.org/10.1016/j.jpubeco.2014.05.002>
- Dwenger, N., Kleven, H., Rasul, I., & Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy*, 8, 203–232.
- European Education and Culture Executive Agency (Eurydice), 2020. The Bologna Process and the European Higher Education Area in 2020. Education and Training - European Commission. URL https://ec.europa.eu/education/policies/higher-education/bologna-process-and-european-higher-education-area_en (accessed 10.18.21).
- Fehr, E., & Gächter, S. (1998). Reciprocity and economics: The economic implications of Homo Reciprocans. *European Economic Review*, 42, 845–859. [https://doi.org/10.1016/S0014-2921\(97\)00131-1](https://doi.org/10.1016/S0014-2921(97)00131-1)
- Ferraro, P. J., Miranda, J. J., & Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. *American Economic Review*, 101, 318–322. <https://doi.org/10.1257/aer.101.3.318>
- Frey, B. S., & Meier, S. (2004). Social comparisons and pro-social behavior: Testing 'Conditional cooperation' in a field experiment. *The American Economic Review*, 94, 1717–1722.
- Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 102, 33–48. <https://doi.org/10.1017/S000305540808009X>
- Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148, 14–31. <https://doi.org/10.1016/j.jpubeco.2017.02.003>
- Harrison, G. W., & List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, 42, 1009–1055. <https://doi.org/10.1257/0022051043004577>
- Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. *Journal of applied psychology*, 87(2), 268.
- Hellerstein, J. K., & Neumark, D. (2008). Workplace segregation in the United States: Race, ethnicity and skill. *Review of Economics and Statistics*, 90(3), 459–477.
- Ito, K., Ida, T., & Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, 10, 240–267. <https://doi.org/10.1257/pol.20160093>

- Lacetera, N., Macis, M., & Slonim, R. (2012). Will there be blood? Incentives and displacement effects in pro-social behavior. *American Economic Journal: Economic Policy*, 4, 186–223.
- Landry, C. E., Lange, A., List, J. A., Price, M. K., & Rupp, N. G. (2010). Is a donor in hand better than two in the bush? Evidence from a natural field experiment. *American Economic Review*, 100, 958–983. <https://doi.org/10.1257/aer.100.3.958>
- Loukkola, T., & Zhang, T. (2010). Examining quality culture. 1, 1. *Europ. Univ. Assoc.*
- Moffatt, P. G. (2021). Experimentics: a survey. *Foundations and Trends® in Econometrics*, 11(1–2), 1–152.
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122, 1067–1101.
- Robertson-Smith, G., 2009. Employee engagement: a review of current thinking. Institute for Employment Studies.
- Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, 119(540), 1422–1439.
- Sundaray, B. K. (2011). Employee engagement: a driver of organizational effectiveness. *European Journal of Business and Management*, 3(8), 53–59.
- Sunstein, C. (2019). *How Change Happens*. MIT Press.
- Thaler, R., & Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.
- van Doorslaer, E., & Koolman, X. (2004). Explaining the differences in income-related health inequalities across European countries. *Health Economics*, 13, 609–628.
- Volpp, K. G., Glick, H. A., Galvin, R., DeGuzman, J., & Audrain-McGovern, J. (2009). A randomized, controlled trial of financial incentives for smoking cessation. *The New England Journal of Medicine*, 11.