Anduiza Perea, Eva; Galais, Carol. «Answering without reading : IMC and Strong Satisficing in Online Surveys». International journal of public opinion research, Vol. 29 Núm. 3 (2017), p. 497-519. DOI 10.1093/ijpor/edw007

Answering without Reading: Instructional Manipulation Checks and Strong Satisficing in Online Surveys[i]

Paper accepted for publication in International Journal of Public Opinion Research (Feb. 2016)

Eva Anduiza . Universitat Autònoma de Barcelona.
eva.anduiza@uab.cat
Carol Galais. Universitat Oberta de Catalunya
cgalais@uoc.edu

Abstract

Some respondents of online surveys click responses at random. Screeners or instructional manipulation checks (IMC) have become customary for identifying this strong form of satisficing. This research first analyzes the factors that condition IMC failures using an online panel survey carried out in Spain (2011–2015). Our data show that the probability of passing a screener depends mainly on the screener's difficulty, the individuals' intrinsic motivations for answering the survey, and past failures. We then address the substantive consequences of omitting those who fail to pass IMCs. We find that this strategy introduces an additional source of bias in descriptive analyses. The paper ends with a discussion of the implications that these findings have for the use of IMCs.

.

Keywords: online surveys, survey quality, survey methodology, disengaged respondents, satisficing, IMC, screeners

1. Introduction

One of the specific problems of online surveys is that their format and mode of administration make it possible to have extreme forms of satisficing. Online surveys are self-administered without supervision from an interviewer. Respondents can rush through the questionnaire without paying attention to what they are asked—failing not only to retrieve from their memories the evaluations, orientations, and facts that are sought out by the researcher, but to even read the question.

The emerging literature on the consequences of satisficing in online surveys uses what are known as screeners or instructional manipulation checks (IMC) to detect respondents' level of disengagement for survey research. Most previous works have analyzed how the failure to pass IMCs is related to poor data quality. Some research has focused on the predictors of failing an IMC, striving to answer why people might pass or fail screeners. Among these works, most pay attention to individual characteristics (such as ability), but little attention has been paid to the role of the IMC's difficulty. Even less attention has been paid to analyzing the implications of failing an IMC for substantive research, carried out only with regard to experimental research designs: what should we do with respondents who fail IMCs and what are the consequences of the different strategies for dealing with these respondents?

In order to contribute to these current debates, we analyze IMCs that were introduced in a six-wave online panel survey carried out in Spain between 2011 and 2015. This paper makes two significant contributions to the literature on satisficing in online surveys. First, it sheds light on the mechanisms that explain the failure to pass IMCs. Past failures and intrinsic motivations are found to play a significant role, but the lion's share is reserved for the IMC's difficulty level. Interestingly, material motivations do not seem to have any effect on the likelihood of failing an IMC. Second, the paper shows that removing people who fail IMCs from the analysis (a strategy that has been suggested for experimental research designs) may widen biases already present due to the sample design and recruiting methods of online surveys.

The paper is organized into four sections. First, we present our expectations about the factors that condition satisficing behaviour in online surveys and the implications of excluding satisficers from the analysis. Second, we describe our IMC and data. Third, we analyze the individual predictors of failing to correctly answer screener questions. Fourth, we assess the implications of excluding satisficers from the analysis. Finally, we discuss our results.

2. Strong satisficing in online surveys

The Problem

Online surveys present a number of specific advantages and disadvantages when it comes to measurement error. Some research suggests that, since there are no social cues given by interviewers, online respondents may be more likely to provide honest answers, hence reducing bias due to social desirability (Comley, 2003; Duffy, Smith, Terhanian & Bremer, 2005). Likewise, online surveys do not have to deal with errors introduced by interviewers (Biemer &Stokes, 1989; Schraepler &Wagner, 2005). However, online surveys may be more vulnerable to other risks. Without the facilitating role of the interviewer, each individual respondent has to read and interpret each question on his or her own, which requires higher levels of cognitive effort and motivation than listening to the interviewer reading a question.

This problem of answering without reading can be considered a case of satisficing. The term “satisficing” is made up of the words “satisfy” and “suffice,” and denotes meeting the minimum criteria for adequacy instead of using optimal procedures. The term was initially used by Herbert Simon to label the mechanisms operating in the decision-making process

(1956). It was later adopted to define a cognitive shortcut taken in the process of answering surveys (Krosnick, 1991). Satisficing consists of giving a sufficiently good answer (that is, a verisimilar, reasonable judgment) when asked a question, but skipping or disregarding some of the steps involved in the optimal answering procedure.

According to the theory of optimal answering, the process of response has four steps. First, respondents are supposed to interpret the meaning of the question. Second, they should recall all the relevant facts and evaluations related to the question. Third, they should integrate and summarize the information necessary to answer the question. Fourth, they should report a summary and an accurate answer (Strack, Schwarz & Wänke, 1991; Tourangeau & Rasinski, 1988; Tourangeau, Rips, & Rasinski, 2000). Weak satisficing involves skipping this very last step, incurring an acquiescence bias, having a tendency to choose the first option given (primacy effects), or selecting non-opinion response options (abstention, "don't know," neglecting to answer, etc.). Strong satisficing happens when the processes of retrieving information from the memory and integrating the information are avoided. This may imply endorsing the status quo instead of change options, failing to differentiate in ratings, or even selecting random answers. Online self-administered surveys provide the possibility of a yet stronger type of satisficing: respondents can skip the first step of the response process as well, by failing to read and process the meaning of the question, and instead jumping directly to a random answer.

Previous research using eye-tracking techniques has found that people indeed use this type of cognitive shortcut in online surveys. Respondents pay only partial attention to the information provided in the survey (Galesic, Tourangeau, Couper & Conrad, 2008). Oppenheimer, Meyvis and Davidenko (2009); Kapelner and Chandler (2010), and Berinsky, Margolis and Sances (2014) all find significant amounts of this kind of disengaged behaviour in online survey responses.

Failing to pass IMCs

Online surveys provide the opportunity for a respondent to answer without reading, but also provide the possibility for researchers to detect this behaviour. To identify those individuals who do not read the instructions or the questions of a survey as carefully as they should, researchers can include in their surveys IMCs or screeners (Oppenheimer et al., 2009). Screeners have been used for a long time in the fields of journalistic interviewing and oral trial proceedings, as well as in the protocols for polygraphs used to detect random responses (Nye and Short, 1957). They belong to the group of techniques aimed at detecting poorly engaged respondents.

Previous work on IMCs has focused mostly on how the failure to pass an IMC in an online survey relates to other indicators of poor respondents' behavior, such as irrational responses, lack of attention, inconsistent answers, or time spent in completion of the survey (Jones, House & Gao, 2015; Berinsky, Margolis, & Sances,2014 Oppenheimer, Meyvis,& Davidenko, 2009; Miller, 2009). This follows earlier research on disengaged respondents, which shows that these respondents rush to complete the survey, use fewer than the average number of words to answer open-ended questions, fall into inconsistencies when answering factual questions, or draw "straight lines"—that is, they give the same answer for a large number of consecutive items (Herzog & Bachman, 1981; Malhotra, 2008). We can conclude from this work that failing to pass an IMC is indeed an indicator of poorer data quality.

A number of issues beyond the relationship between IMC failures and data quality, however, require further attention. We shall distinguish two broad questions to which this paper seeks to make a contribution: the causes that lead to IMC failure and the consequences of different ways of dealing with those respondents who failed the IMC for substantive research.

First, we need more research on the factors that influence the probability of failing an IMC. The antecedents of this research strand are rooted in satisficing theory, which suggests that suboptimal responses to survey questions are boosted by individuals' characteristics, such as (lack of) ability and motivation (Bishop & Smith, 2001; Krosnick & Alwin, 1987; Krosnick, Narayan & Smith, 1996; Narayan &Krosnick, 1996). Motivations for conscientiously answering an online survey might be material or intrinsic. As for the first set of motivations, respondents may want to rush through the survey to collect the incentive at the end, as has been suggested (Downes-Le Guin, 2005; Berinski et al., 2014; Zagorsky & Rothon, 2008). However, respondents who answer online surveys for monetary or material compensation may also wish to be contacted again for the same reasons, and hence may want to do a more careful job when completing the survey. Intrinsic motivations, in turn, are related to the interest that the person has in the topic of the survey. In a survey on political attitudes and behaviour, respondents with lower levels of interest in politics or political engagement are expected to be more likely to become uninterested and rush through the survey without paying much attention to the questions.

A few works have analyzed the differences between those respondents who fail and those who pass IMCs, yielding mixed results or findings that are not always consistent across different studies or with the postulates of satisficing theory. Kapelner and Chandler (2010) find that this behavior is less frequent among women, and it decreases with age and education (but not with the need for cognition), and with some indicators of motivation (e.g., number of words in feedback). Nevertheless, they also find that failures were unrelated to self-reported motivation. Oppenheimer et al. (2009) did not find any effects of age, gender, self-reported motivation, or material motivation on failing IMCs (except for the need for cognition).As they argue themselves, this lack of differences may be due to the characteristics of their samples (small groups of students) and things may work differently in larger and more heterogeneous samples. Knowing to what extent failing to answer a screener question is conditioned by age, education or motivations is important in itself, but it also has relevant implications for how to deal with those respondents who fail an IMC, as we shall see later on.

Furthermore, there is some debate on whether respondents can be clearly divided into attentive and inattentive categories. Barge and Gehlbach (2011) estimate that “disengaged respondents” are not a marginal category of individuals, as a large majority of respondents relax their engagement at some point during a questionnaire—a problem that seems to be more significant in online surveys (Chen, 2011; Fang, Wen & Prybutok, 2014).Respondents may even choose to provide attentive responses to other questions but not in the IMC (Jones et al., 2015). It seems that, in any case, attention varies with time, and hence although respondents may be inattentive sometimes, they may become more attentive later on (as noted by Berinsky et al., 2014). This means that past failure to pass an IMC does not necessarily imply future satisficing. It is hence important to know to what extent past IMC failures condition the probability of failing again, as this can help to assess to what extent IMCs can detect a punctual or a permanent situation of poor respondents’ attention. In other words, the relationship between past and subsequent IMC failures gives us an approximate idea of how badly respondents who fail screeners can affect the quality of our data.

One last under-researched aspect that might help us understand why some respondents fail to pass IMCs is the type of screener used. According to previous research, the proportion of people who fail to pass IMCs varies quite a lot (from a few percentage points to almost half the sample; see Oppenheimer et al., 2009), and this may be at least partly due to the design of the screener. Berinsky et al. (2014) compare on-topic (harder) with off-topic (easier) screeners. The length of the screener may also have consequences for the rate of passage. Indeed, according to satisficing theory, task difficulty boosts respondents’ mindlessness (Krosnick & Alwin, 1987; Krosnick, Narayan & Smith, 1996). Therefore, short, easy screeners may let moderately inattentive people pass and detect only severe cases of lack of attention. On the other hand, long screeners may induce failure among partially attentive people; that is, they may induce some level of satisficing. Comparing different IMCs with different levels of complexity may help to shed some light on the extent to which the design of the IMC may condition the number of people who fail these questions as well as the characteristics of those who fail.

The second fundamental question that requires further research is how to deal with this kind of respondents' behaviour, and at what price. According to Baker et al., "there is nothing in the research literature to help us understand how significantly any of these respondent behaviours [failing IMC] affect estimates" (2010:47). Hence, how shall we deal with respondents who fail to pass IMCs when trying to carry out substantive research, such as describing or explaining political behaviours or political attitudes? What are researchers supposed to do with those who fail IMCs?

The literature suggests providing training for respondents (which would increase attention and reduce noise; see, for instance, Oppenheimer et al., 2009). But outside the lab environment, the effects of training on noise reduction are not always found, and the costs in terms of dropouts seem to be high (see Berinsky et al., 2014). Oppenheimer et al. (2009) argue that, provided that failures are related to impaired performance, and that those who fail are not too different from those who do not fail, the former should be removed from the analysis so as to increase statistical power in experimental research designs. However, if those who pass IMCs are different than those who do not pass them—in terms of characteristics that are related to outcomes of interest—then inclusion or exclusion of these respondents may have consequences for our estimators both when describing the characteristics of our sample and when analyzing relationships.

This is not trivial. Considering that non-probabilistic online surveys entail coverage error and over-representing a certain type of internet user (more educated, younger, etc.; see Rohde and Shapiro, 2000) and that these kinds of surveys tend to enroll individuals who are more sophisticated and attracted by politics than the rest of the population (Chang and Krosnick, 2009), getting rid of those individuals who fail the screeners would mean dropping those who are more similar to the average citizen, both in terms of socio-demographics and political sophistication. If, as satisficing theory predicts, those who fail have lower levels of ability (for instance, as measured with education) and motivation (such as interest in politics), excluding them would introduce an additional source of bias to the results.

Berinsky et al. (2014) argue that those respondents who fail an IMC should not be excluded from the analysis, as they do not constitute a stable group. Furthermore, they point out that the lack of attention is a real circumstance, and excluding less attentive respondents exacerbates the concern that experiments overestimate treatments because the artificial setting increases attention. The literature has focused on the consequences of keeping/removing individuals who fail IMCs in the context of experimental research designs, but this reasoning could be also applicable to samples recruited for observational studies. Our expectation in this case is that removing individuals who fail to pass IMCs may contribute to aggravated bias. To put it simply, individuals who fail IMCs are different from those who pass them because they are more representative of the general population; excluding them from the sample will exacerbate the biases present in an already over sophisticated sample.

Our hypotheses can then be summarized as follows. We expect IMC failure to behave like other forms of survey satisficing, in that abilities, motivations and the difficulty of the task—probably the most disregarded aspect—should predict IMC failures. We expect IMC failure to be an inconsistent, intermittent behaviour that by no means identifies a "recidivist" disengaged profile. Failing IMCs in the previous wave is expected to influence only to a limited extent the probability of failing the IMC in a subsequent wave. Precisely because we expect some differences between those who pass and those who fail IMCs, excluding those who fail from the analyses is expected to affect the results of the substantive, descriptive and explanatory analyses, making the remaining sample more sophisticated and biasing any estimations accordingly.

3. Research design and data

Our data are taken from a seven-wave online panel survey.[ii] The longitudinal nature of the data allows us to analyze not only the effect of ability and motivations, but also to what extent failures are constant or variable along time and how the same individuals react to different IMCs. The panel consists of Spanish citizens between the ages of 16 and 45 (in wave 1) with internet access. The first wave of the panel does not include any IMC questions, so this wave is not considered in the analysis. Wave 2 included 2,433 respondents, and wave 7 keeps 1014 of them. From wave 2 on, we collected information (9,971 observations equivalent to individual * waves) from 2,518 different individuals.

Respondents, recruited by a survey firm using commercial banners, received incentives for completing the survey, ranging from 22 to 37 points, depending on the length of the questionnaire. The points could then be exchanged for material goods. The expected economic value of each point is 0.1 Euros. The questionnaires had an average duration between 18 and 34 minutes, depending on the wave, and included questions on standard political attitudes and behaviours: media consumption, interest in politics, political confidence, political participation, leader evaluation, satisfaction with the political and economic situation, evaluation of government and opposition, vote intention and recall, etc. We introduced an IMC in each wave beginning with wave 2, at approximately the same point in each questionnaire, after about one-third of the questions. In particular, for waves 2, 5 and 7, the IMC came after a battery on the probability of voting for each party, and in waves 3, 4 and 6 it came after the question of evaluating the role of the parties in the government and opposition.

The specific IMCs used in each wave are reported in Table 1. IMCs included the correct answer within the visible usual response options in waves 2, 3, 4 and 5, with different numbers of response options. In waves 6 and 7, the respondent was instructed not to click on any particular response of the scale, but to click one of the words of the instructions (wave 6) or to pass to the next question (wave 7).

More specifically, in wave 2 the IMC had a 0 to 10 scale as the response option, and respondents were requested to select the number 2 from the scale. Almost 9% of respondents failed to pass this IMC. Waves 3, 4, and 5 included similar ICMs, but with 5 ordered response options. About 5–7% failed to pass these checks. Wave 6 included a longer IMC with about three times as many words as the traps included in waves 3, 4, or 5. The correct answer was not included in the visible response options, and thus nobody could pass the IMC just by chance, without reading the whole question. This type of question is similar to the one devised by Oppenheimer et al. (2009), who find fail rates of almost 50%, or by Kapelner and Chandler (2010), who apply the survey to Amazon Mechanical Turk—a highly motivated and sophisticated sample—and find a fail rate below 20%. In our case, almost 60% of the sample failed to pass the IMC in the 6th wave. Finally, the last wave included a shorter question, where people were presented with a scale of 0 to 10 but were instructed to pass without producing any answer. Note that this option (skipping a question) was seldom allowed in the questionnaire (only for questions on income and vote intention). In this wave, which was answered mostly by the survivors of a 7-wave panel that had been running for over 4 years, only 3% of the sample failed to pass the IMC.[iii]

(Table 1 here)

4. Analysis

Why do people fail IMCs?

To address why people fail IMCs, we run a number of logistic regressions, which are presented in Table 2. In these analyses, the dependent variable is failing to pass an IMC. The models include as explanatory variables ability (education), motivations (material and intrinsic), past behaviour (failure to pass an IMC in the previous wave) and the characteristics of the IMC (length in words). It also controls for age, sex and income.

(Table 2 about here)

All variables are recoded so that 0 represents the lack of or a minimum value of the variable and 1 indicates the presence or maximum value. Education is a four- category variable (primary education or less, secondary education, university degree—set as a reference—and postgraduate studies). Personal income is based on a ten-point scale ranging from less than €300 per month to more than €6,000 per month. As the survey is mainly concerned with politics, intrinsic motivation is measured with party identification (value 1identifying those claiming to have a party allegiance) and interest in politics (based on a four-point scale from not at all to a lot). Material incentives are measured in wave 4 with a dummy which takes value 1when the respondent mentioned the material incentives offered by the survey firm among the two main reasons for completing the questionnaire (other potential reasons offered were "for the possibility to give my opinion", "out of curiosity", "because I am interested in the topic", "because the survey is carried out by the CIS", "because I always answer surveys" or "others").

The extent to which previous IMC failures condition future fails during the study is modeled, including a lagged measure of the dependent variable. Finally, as an indicator of the difficulty of the IMC we take into account the question length as measured by the number of words used in the question (Hox & Borgers, 2001; Holbrook, Cho & Johnson, 2006). Note that this variable is measured at the wave level—or, if preferred, at the IMC level—and therefore its value is identical for all the individuals within each wave.

The number of observations per respondent varies from one to six. In other words, some individuals have repeated observations, which might lead to the underestimation of the error terms and violate the assumption of regression analysis that they are independent. Hence, we have clustered the estimations' standard errors by individual, a strategy which allows for arbitrary correlation within individuals and also corrects for autocorrelation (Rogers 1993). This also yields a large N (9,351 observations for the dependent variable) and guarantees robust inference.[iv]

Table 2 includes six different estimations of the likelihood of failing an IMC. Model 1 is the baseline, and includes the individual variables referring to sociodemographics, ability and motivations. Gender, age or income do not exert a significant effect on the probabilities of failing a screener. Education does have a significant effect: both those with the minimum level of education and, interestingly, those with the maximum level, exhibit a higher propensity to fail IMCs than those with college education (the reference category). The two variables tapping into intrinsic motivations have significant, negative effects: the more interested the respondent is in politics, the less likely she is to fail the IMCs. Likewise, partisanship reduces failures. Acknowledging a material interest in taking the survey does not affect the probabilities of failing the screener.

Model 2 includes the same variables as model 1 with a lagged measure of the dependent variable. This variable has, as expected, a significant, positive effect, meaning that those respondents who failed a screener in a previous wave of the study are more likely to fail it again in the following wave. The effect is not particularly strong when compared, for instance, with interest in politics. The observed effects in the previous model for the intrinsic motivations hold when this lag is considered.

Because almost half of the sample fails the IMC of the 6th wave , we replicate the analysis of model 2 without the 6th wave in model 3.This is meant to show to what extent those cases of extremely strong satisficing (failing very short, and therefore easier, IMCs) influence future extreme failures as well. We would expect the likelihood of failing an IMC to be more conditioned in model 3 (where very few people fail the screeners) than in model 2. We can see that the effect is indeed stronger, but when we consider average marginal effects the lagged variable increases the likelihood of failing the IMC by 6% in model 2 and 9% in model 3, so the difference is not that great. Hence, even in the case of very extreme satisficing, past satisficing only has a limited effect. Education and age significantly affect the likelihood of extreme satisficing, and the effect of intrinsic motivations increases. The model fit is also better in this model, suggesting that failure to pass shorter IMCs conforms better to the satisficing theory than failure to pass longer IMCs. Model 4 replicates the same model only for wave 6, which includes the longest IMC, which more than half of the sample failed. The only significant predictors of the likelihood of failing to pass the IMCs in this wave are interest in politics and lagged fails. Although this model is fairly similar to the former, the possibility exists that long IMCs induce satisficing among average respondents.

Model 5 incorporates ability, motivations and control, and adds a measure of IMC difficulty: the length of the question. As with all other variables, this indicator has been rescaled to range between 0 and 1. Therefore its coefficient should be interpreted as the effect of going from the minimum to the maximum number of words used in the IMCs included in the panel survey. IMC length has a positive, significant effect. Model 6 includes all the explanatory variables. We see that the negative effects of ability, intrinsic motivations, past failures and IMC difficulty hold. IMC difficulty, previous fails and interest in politics appear as the most relevant explanatory factors in this model.

(Figure 1 about here)

Figure 1 represents the average marginal effects that our last estimation yields. Average marginal effects (AME) are changes in the propensity to fail an IMC due to one predictor while keeping all the other variables included in the model at their observed values. This allows us to compare the contribution of each these predictors to our understanding of IMC failing. We see that among our controls, the only variable that helps to predict failure is age. Variation from its minimum to its maximum value, with all else being kept equal, decreases the chances to fail by almost 12%. Individuals with no studies or only very basic ones are 4% more likely to fail an IMC than those with some college education. Previous failures increase the likelihood of failing again by 11%, and the two indicators of intrinsic motivations, interest and party identification, decrease the chances of failing by 11% and 3%, respectively. But the greatest effect is for the variable tapping into the IMC difficulty. The longest IMC question has a 28% greater chance of being failed than does the shortest IMC question. Alternatively, we can interpret this finding by stating that every additional word represents, on average, a 0.5% increase in the probability of failing a screener, although we must take this interpretation with a pinch of salt. Indeed, we are aware that the difficulty of the task associated to this IMC is not only due to its length.

Excluding or keeping individuals that fail to pass an IMC?

As stated by Oppenheimer (2009), excluding those who fail an IMC from a database might only be advisable when those who fail and those who pass are relatively similar. But we have seen that those who fail tend to be younger, less educated, and less motivated by political issues than those who pass. Hence, omitting those who fail IMCs might aggravate the typical sample biases that plague non-probabilistic surveys. All in all, deleting the information given by respondents who failed to pass an IMC may increase the biases already present in some samples.

In order to address the question of what to do with those who fail IMCs and what the implications for different courses of action are, we have adopted a benchmarking strategy. We have selected a high-quality survey which is representative of the Spanish population. The 2973 Centro de Investigaciones Sociológicas' (CIS) study was conducted in December 2012, approximately at the mid-point of our panel survey. After clearing the dataset of all those who do not regularly use the internet or who do not fall within the age span covered by our online panel, we carry out different strategies of dealing with our respondents who fail to pass IMCs and see how these results compare to the benchmark in different descriptive and explanatory substantive analyses.

More specifically, we consider four different strategies:

1. Keeping all respondents who fail the IMCs in the analyses using the whole online panel sample. This strategy considers all information provided by respondents, even if it is coming from someone who has failed an IMC.
2. Dropping the observations in a wave when the individual fails the IMCs in that wave, but keeping the observations of that same respondent in waves where she passes the IMCs. This strategy considers responses as poor only if they are given in a wave when the respondent failed the IMC.
3. Dropping all observations for individuals who have failed two or more of the six IMCs. This strategy considers responses as poor if they come from individuals who have failed to pass an IMC in two or more waves, even if these individuals have passed other IMCs in other waves.
4. Dropping all the individuals who failed any IMC in any wave. This strategy excludes any information coming from any individual who has ever failed an IMC within the panel.

As for the variables concerned, we consider two very popular variables in public opinion and political behaviour studies: political interest and a scale of non-electoral political participation. Interest in politics is measured using the standard question "To what extent would you say you are interested in politics? Not at all/ A little/ Quite / A lot". Political participation is calculated as an additive scale using 5 indicators: petition, boycott, donation, demonstration and strikes. Respondents were asked if they had engaged in each of these modes of participation in the past (in the previous 12 months for the CIS survey, or in the previous 6 months for the panel study). Both variables—as well as their predictors of gender, age and interest—have been recoded so as to range between 0 and 1.

As a first descriptive analysis we simply compute the mean value of interest in politics and of an index of political participation for the benchmark and for the different strategies. Figures 2 and 3 represent the deviation introduced by these strategies with respect to the CIS benchmark.

(Figures 2 and 3 about here)

The benchmark shows an average level of interest of 0.39. All the possible strategies regarding those who fail IMCs with the online panel fall well above this benchmark, showing that our panel overestimates the degree of interest in politics. This is something we already know, as we have a sample that over-represents highly educated people who have survived a 7-wave political attitudes panel. However, the first strategy that includes all respondents gets the closer estimate. Any other strategy aggravates the bias already present in our sample. The approach of dropping all the observations from an individual who has failed any screener yields the most biased estimation of the interest in politics. The same is valid for political participation: the whole sample gets the closer estimation and the worst estimate is obtained when computed from the sample that excludes those who failed any IMC.

(Table 3 about here)

For the explanatory analysis we predict political participation using interest in politics and controlling for gender, age and education, as shown in Table 3.We know that interest in politics is an important predictor of political participation, hence we expect a strong association between the two variables. The highest R-squared value—outside the benchmark—is observed for the complete sample, without excluding observations on the basis of their IMC failures. The effect of interest in politics is very similar in all the alternative samples to the effect observed in the benchmark, but is somewhat closer in the second strategy (third model), where the observations for those who failed an IMC in a panel wave have been excluded only for that wave. On the other hand, the closest effect for sex as compared to the benchmark is observed when applying the tougher strategy; that is, dropping all the individuals who failed any of the six screeners. The effect of age is non-significant in all models. Finally, the effect of education differs to some extent in all the alternative strategies. The lowest level of education has a negative, significant effect in all the models, but the closer estimate to the benchmark is obtained when we exclude the observations of those individuals who failed an IMC for that same wave of the panel, while saving their information in subsequent and previous waves. In the benchmark, a clear negative significant effect is also observed for those with a secondary level of education when compared to those with some tertiary education. This effect is not observed in any of the samples drawn from our online panel. Curiously enough, we observe a positive effect of the highest level of education only when we apply the strategy of deleting the observations of those respondents who failed an IMC within that wave. Therefore, this suggests that the deviations from the benchmark can be different for different variables.

5. Conclusions

Failing to answer IMCs correctly can be explained as a strong form of satisficing. A respondent's ability (education), and particularly her intrinsic motivations (in this case, her level of interest in politics and partisanship) reduce her likelihood of failing IMCs. It is noteworthy that being motivated by the material rewards used as a decoy for completing the survey does not affect the chances of failing the IMC. We can derive from this finding that those respondents recruited by survey firms who are particularly interested in the material incentives do not necessarily provide poor responses that affect the quality of the data. This finding provides additional support for the idea that material incentives are not a problem (see Singer & Ye, 2013), nor are "professional" respondents (Chang & Krosnick, 2009; Hillygus, Jackson &Young, 2014).

Our longitudinal data allowed us to assess whether individuals systematically pass or fail IMCs. Past failures only condition future failures to a limited extent, so it is difficult to talk about generalized “disengaged respondents” or “shirkers”. Attention as measured with IMCs comes and goes from one wave to the next. Our data show that having failed an IMC in the previous wave increases the probability of failing again in the subsequent wave by about 11%, but the marginal effect of this variable is not larger than the one observed for intrinsic motivations.

According to our results, the length of the IMC is an important predictor of the likelihood of failing. The longitudinal structure of the data does not provide the perfect test for the relationship between number of words and propensity to fail (a survey experiment manipulating the length of the IMCs would be a better design), as the difficulty of the task and the length of the question are intertwined and overlap in our wave 6 IMC. Yet our results point at some findings that require further attention. Long IMCs would only let extremely attentive respondents pass, and leave partially attentive respondents as fails. One wonders to what extent a long and difficult IMC may not be inducing some amount of satisficing, and even producing other effects on subsequent questions.

A large part of the research conducted on IMC finds associations with other signs of respondents’ disengagement. However, these findings do not help us in substantive research. What should we do with respondents who do not pass IMCs? Some scholars have suggested filtering them (as this would reduce noise) while others have suggested that it is better to keep them (as filtering them would threaten external validity). We have tested different strategies to deal with these respondents with a benchmark (a representative face-to-face survey). Our results suggest that for descriptive analyses there is a significant risk of increasing bias if we remove those respondents who fail IMCs (using any of the different strategies for removal). For explanatory analysis, however, the results do not seem to change much regardless of the strategy we employ. Small differences seem to depend on the variable considered.

IMCs seem to be a good indicator of respondents' attention level and the quality of their responses. Easy, short IMC questions detect severe cases of lack of attention, while hard, long IMC questions identify the outstandingly attentive respondents. However, it is not clear that we want to get by without either of these two groups. A proper assessment of (varying levels of) attention would require repeated IMCs throughout the questionnaire. This may be feasible for researchers interested in data quality issues, but it is not always possible for researchers who are interested in other substantive issues, where questions on public opinion, attitudes and behaviours may take priority. In the absence of such possibilities for observational analysis, we may be better off considering all respondents in the analysis. Since repeated IMCs may also have unintended consequences among respondents (introducing discomfort and uneasiness), we may even be better off without any IMCs at all.

References

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., ... & Zahs, D. (2010). Research synthesis AAPOR report on online panels. *Public Opinion Quarterly*, *74*(4), 711-781.

Barge, S., & Gehlbach, H. (2011). Using the Theory of Satisficing to Evaluate the Quality of Survey Data. Research in Higher Education, 53(2), 182–200. doi:10.1007/s11162-011-9251-2

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys. *American Journal of Political Science*58 (3), 739-573.

Biemer, P. P., & Stokes, S. L. (1989). The optimal design of quality control samples to detect interviewer cheating. *Journal of Official Statistics*, *5*(1), 23-39.

Bishop, G., & Smith, A. (2001). Response-order effects and the early Gallup split-ballots. Public Opinion Quarterly, 65(4), 479-505.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the internet comparing sample representativeness and response quality. *Public Opinion Quarterly*, *73*(4), 641-678.

Chen, P.-S. D. (2011). Finding Quality Responses: The Problem of Low-Quality Survey Responses and Its Impact on Accountability Measures. Research in Higher Education, 52(7), 659–674. doi:10.1007/s11162-011-9217-4

Comley, P. (2003) Innovation in online research – who needs online panels?, MRS Research Conference Paper (Vol. 36, pp. 615-639).

Downes-Le Guin, T. (2005, June). Satisficing behavior in online panelists. In*MRA Annual Conference & Symposium, June*.

Duffy, B., Smith, K., Terhanian, G., & Bremer, J. (2005). Comparing data from online and face-to-face surveys. *International Journal of Market Research*,*47*(6), 615.

Fang, J., Wen, C., & Prybutok, V. (2014). An assessment of equivalence between paper and social media surveys: The role of social desirability and satisficing. Computers in Human Behavior, 30, 335–343. doi:10.1016/j.chb.2013.09.019

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-Tracking Data New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding. Public opinion quarterly, 72(5), 892-913.

Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*(4), 549-559.

Hillygus, D. S., Jackson, N., & Young, M. (2014). *Professional respondents in non-probability online panels*. In Mario Callegaro, Reg Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick and Paul J. Lavrakas (eds.) *Online panel research: a data quality perspective*. Wiley, Chichester, 219-237.

Holbrook, Allyson, Young Ik Cho, and Timothy Johnson(2006). "The impact of question and respondent characteristics on comprehension and mapping difficulties." *Public Opinion Quarterly* 70(4): 565-595.

Hox, J. J., and N. Borgers (2001). "Item nonresponse in questionnaire research with children*." Journal of Official Statistics* 17(2):321-335.

Jones, M. S., House, L. A., & Gao, Z. (2015). Attribute Non-Attendance and Satisficing Behavior in Online Choice Experiments. *Proceedings in Food System Dynamics*, 415-432.

Kapelner, A., & Chandler, D. (2010). Preventing Satisficing in online surveys. In*Proceedings of*.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, *5*(3), 213-236.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3), 213-236.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. Public Opinion Quarterly, 51(2), 201-219.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. New Directions for Evaluation, 1996(70), 29-44.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New directions for evaluation*, *1996*(70), 29-44.

Malhotra, N. (2008). Completion Time and Response Order Effects in Web Surveys. Public Opinion Quarterly, 72(5), 914–934. doi:10.1093/poq/nfn050

Miller, C. (2009) Respondent technology preferences. CASRO Technology Conference, New York

Nye, F. I., & Short, J. F. (1957). Scaling delinquent behavior. *American Sociological Review*, *22*(3), 326-331.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872.

Rogers, W. H. (1993). "Regression Standard Errors in Clustered Samples." *Stata Technical Bulletin* 13:19-23.

Rohde, Gregory L., and Robert Shapiro.(2000). *Falling Through the Net: Toward Digital Inclusion*. Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration and National Telecommunications and Information Administration.

Schraepler, J. P., & Wagner, G. G. (2005). Characteristics and impact of faked interviews in surveys–An analysis of genuine fakes in the raw data of SOEP.*Allgemeines Statistisches Archiv*, *89*(1), 7-20.

Simon, Herbert A. (1956) "Rational choice and the structure of the environment", *Psychological Review*, vol. 63, March, compiled in, and quoted from, de, Simon (1957: 261-273).

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(1), 112-141.

Strack, F., Schwarz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition*, *9*(1), 111-125.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Tourangeau, Roger, and Kenneth A. Rasinski. "Cognitive processes underlying context effects in attitude measurement." *Psychological bulletin* 103.3 (1988): 299.

Zagorsky, J. L., & Rhoton, P. (2008). The effects of promised monetary incentives on attrition in a long-term panel survey. *Public opinion quarterly*,*72*(3), 502-513.

Table 1

IMC questions by wave

Wording	Responses	Wave	N failures (%)	N correct (%)	Total (%)
In order to verify that the browser works properly and that we are collecting all your answers, could you please select the number two on the following scale?	Scale 0–10	2. April 2011	206 (8.5%)	2,227 (91.5%)	2,433 (100%)
In order to verify that the browser works properly and that we are collecting all your answers, could you please select the category bad from the list below?	Very good Good Fair Bad Very bad	3. October 2011	134 (6.8%)	1,845 (93.2%)	1,979 (100%)
In order to verify that the browser works properly and that we are collecting all your answers, could you please select the category fair from the list below?	Very good Good Fair Bad Very bad	4. April 2012	79 (4.6%)	1,638 (95.4%)	1,717 (100%)
Could you select the category fair from the list below in order to verify that the browser works properly and we are collecting all your answers?	Very good Good Fair Bad Very bad	5. April 2013	113 (6.4%)	1,644 (93.6%)	1,757 (100%)
People are currently very busy and have no time to be informed about government decisions. However, although some people pay attention to issues related to politics, not everybody reads all questions carefully. To show us that you are really reading this, please ignore the question we are asking you next and click over the word 'decisions' in the next sentence. How do you follow information regarding government political decisions?	With a lot of interest With some interest With little interest With no interest (Requested answer is not one of these)	6. May 2014	633 (59.1%)	438 (4.9%)	1,071 (100%)
In order to verify that the browser works properly and that we are collecting all your answers, could you please go to the next question without answering	Scale 0–10 (Requested answer is not one of these)	7. April/May 2015	34 (3.3%)	980 (96.7%)	1014 (100%)

Note: all the screeners where introduced at the end of the first third of each questionnaire, after the questions addressing political interest, media consumption, political particuation, national identities political efficacy, nationalism and other core political attitudes and values.

Table 2.

Logistic regressions of failing to pass the IMC

	(1) baseline	(2) Lagged fail	(3) Lagged fail without W6	(4) Lagged fail only W6	(5) Length IMC	(6) Length IMC + lagged fail
Woman	-.13	-.12	.10	-.06	-.07	-.01
	(.08)	(.08)	(.15)	(.15)	(.11)	(.11)
Age	-.16	-.07	-1.97***	-.37	-1.68***	-1.62***
	(.24)	(.22)	(.45)	(.43)	(.32)	(.32)
Income	-.02	-.07	.25	.42	.41	.33
	(.23)	(.24)	(.44)	(.38)	(.28)	(.30)
Education:						
primary studies or less	.20	.21	.71***	.16	.52***	.50**
	(.11)	(.11)	(.21)	(.22)	(.15)	(.16)
secondary studies, HS or equivalent	.12	.07	.24	-.11	.19	.088
	(.10)	(.10)	(.17)	(.19)	(.13)	(.13)
tertiary education, college or equivalent (ref.)	-	-	-	-	-	-
PhD & Master	.38*	.39**	.33	.11	.26	.25
	(.15)	(.14)	(.31)	(.24)	(.19)	(.18)
Interest in politics	-1.26***	-1.16***	-1.81***	-1.01***	-1.60***	-1.47***
	(.16)	(.16)	(.35)	(.29)	(.21)	(.22)
Party identification	-.42***	-.37***	-.78***	.09	-.47***	-.42***
	(.08)	(.09)	(.17)	(.17)	(.11)	(.12)
Material motivation	-.03	-.00	-.04	.06	-.04	.01
	(.08)	(.08)	(.16)	(.16)	(.12)	(.11)
Lagged fail		.55***	2.23***	1.94***		1.54***
		(.11)	(.19)	(.53)		(.14)
Length of IMC					3.54***	3.96***
					(.11)	(.32)
Constant	-1.15***	-1.22***	-1.62***	.85**	-1.49***	-2.05***
	(.16)	(.16)	(.29)	(.31)	(.21)	(.23)
Pseudo R-Squared	.032	.037	.191	.04	.297	.37
Obs.	7192	5383	4585	798	7192	5383

Standard errors in parentheses
* p<.05, ** p<.01, *** p<.001

Table 3.

OLS estimation of political participation following different strategies with IMCs fails.

	(1) Bench-mark	(2) Online panel	(3) Omitting those who fail within each wave	(4) Omitting individuals failing any IMC	(5) Omitting those who fail more than 2 IMC
	b/se	b/se	b/se	b/se	b/se
Interest in politics	.308***	.317***	.307***	.311***	.306***
	(.025)	(.014)	(.016)	(.017)	(.015)
female	.030*	.018*	.018	.024*	.018*
	(.015)	(.009)	(.010)	(.010)	(.009)
age	-.018	-.009	.004	.004	-.010
	(.029)	(.022)	(.026)	(.027)	(.023)
Education:					
primary studies or less	-.129**	-.036**	-.048***	-.041**	-.039***
	(.040)	(.011)	(.013)	(.014)	(.012)
secondary studies, HS or equivalent	-.091***	-.009	-.011	-.019	-.008
	(.016)	(.010)	(.011)	(.012)	(.010)
tertiary education, college or equivalent (ref.)	(.)	(.)	(.)	(.)	(.)
PhD & Master	.021	.025	.030	-.002	.021
	(.034)	(.016)	(.017)	(.018)	(.016)
R-Squared	.112	.104	.098	.096	.096
N	1730	12071	8220	7826	11162

Constant not shown. Standard errors in parentheses
* p<.05, ** p<.01, *** p<.001

Note: CIS respondents over 50 years old and those who have not used internet for the last 3 months have been dropped for the analyses

The scale values have been recoded so as to range from 0 (minimum participation) to 1 (maximum participation). Non-dychotomous independent variables (Interest in politics, age) have also been recoded so as to range between 0 and 1.

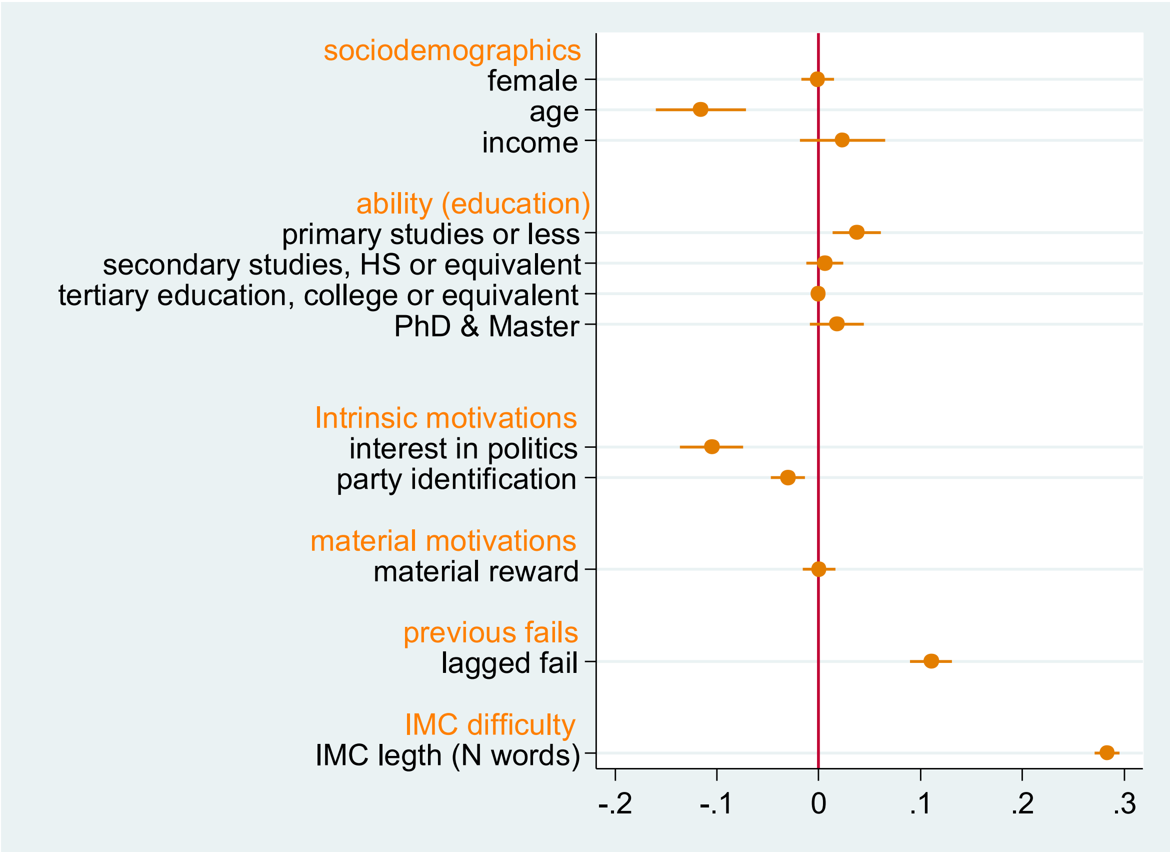

Figure 1.
Average marginal effects for the probabilities of failing to pass the IMC (from table 2, model 6, 95% CI)

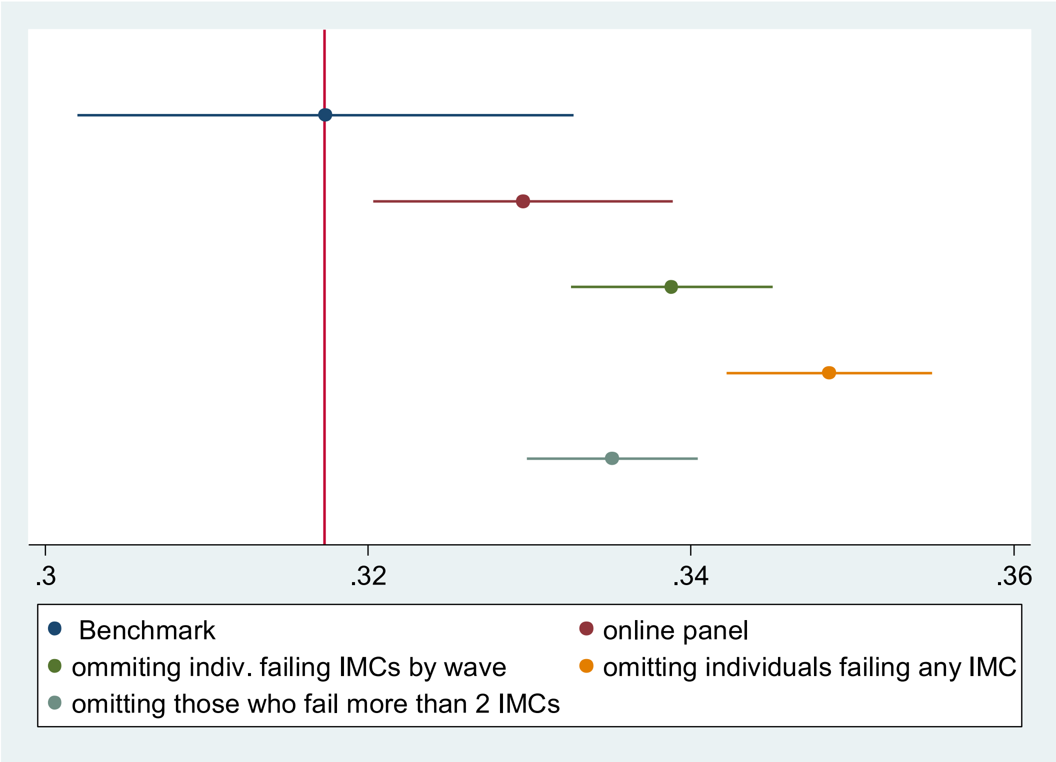

Figure 2: Mean political participation (95% CI)for benchmark and different strategies.

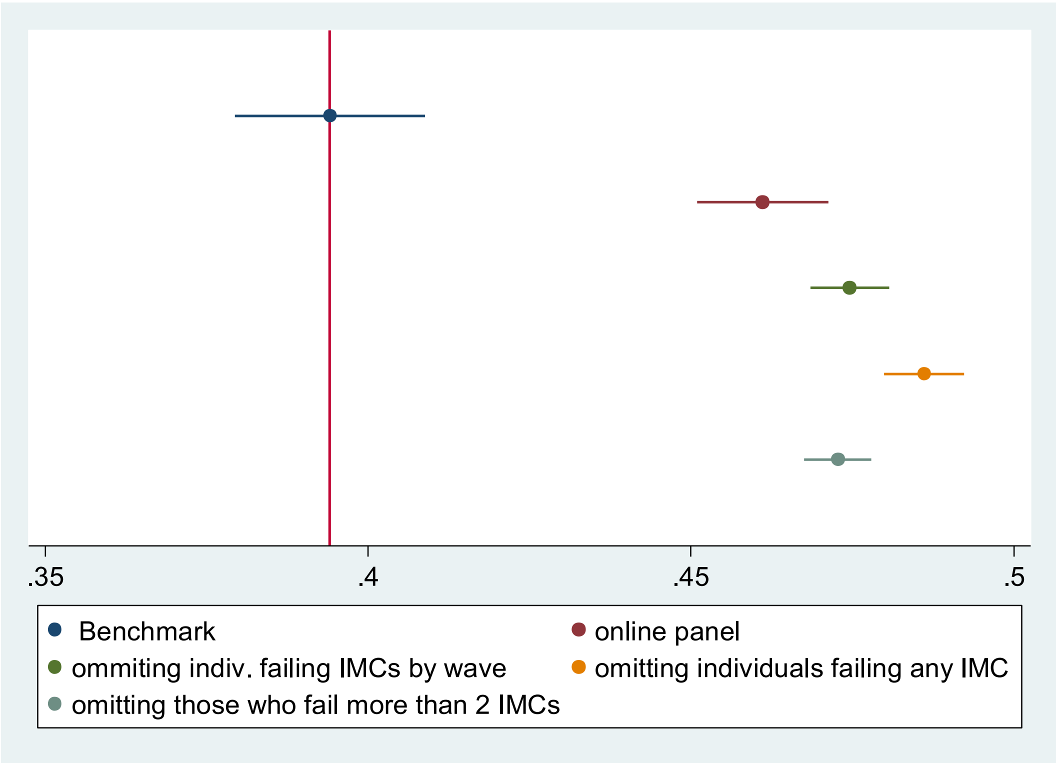

Figure 3: Mean interest in policis (95% CI) for benchmark and different strategies

APPENDIX

Table A1. Survey sample

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
Total N	2433	1979	1717	1757[a]	1071	1014
Fieldwork dates	May 11- May 25, 2011	November 9- November 18, 2011.	May 11- May 30, 2012.	May 17- June 4, 2013 and October 16- 27 2013.	May 5-12 2014	April 27- May 8 2015

Source: Our elaboration on CIS 2855

[a]: this wave recovered some indidivuals that had participated in the study but withrawed it in previous waves.

Table A2. Descriptives of the main variables per wave

Variable	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Total
Female	.50	.49	.48	.49	.47	.48	.49
	(.50)	(.5)	(.50)	(.5)	(.5)	(.5)	(.5)
Age	32.1	32.4	32.7	34.6	35.7	37.03	33.05
	(7.3)	(7.05)	(6.9)	(7.2)	(7.14)	(7.1)	(7.42)
Interest pol.	1.337	1.35	1.4	1.37	1.3	1.4	1.38
	(.83)	(.86)	(.85)	(.89)	(.86)	(.87)	(.85)
Income	3.86	3.79	3.66	3.71	3.74	3.83	3.8
	(1.85)	(1.81)	(1.82)	(1.89)	(1.94)	(1.83)	(1.86)
Education (4 cat)	2.27	2.35	2.37	2.44	2.49	2.5	2.4
	(.95)	(.92)	(.92)	(.91)	(.92)	(.92)	(.9)
Party identification	.67	.7	.67	.56	.62	.68	.67
	(.46)	(.46)	(.47)	(.5)	(.49)	(.47)	(.47)
Response reward	.34	.34	34	.35	.34	.34	.34
	(.47)	(.47)	(.47)	(.48)	(.48)	(.47)	(.47)
IMC (failures)	.074	.067	.046	.066	.59	.034	.12
	(.27)	(.25)	(.21)	(.25)	(.49)	(.18)	(.33)
IMC: (N words)	27	27	26	24	87	25	32.4
	(0)	(0)	(0)	(0)	(0)	(0)	(18.9)
Lagged IMC	0	.07	.06	.04	.06	.61	.123
		(.26)	(.25)	(.2)	(.24)	(.49)	(.33)

Cell entries are variable means. Standard deviations in parentheses.

The table reflect the original values of the variables before being recoded so as to range between 0 and 1.

The material motivations indicator was only asked once, in wave 5. The values of this variable for each individual have been carried forward and backward so as it is constant within each individual in the estimations displayed in Table 2.

[i] Authorship is alphabetical to reflect the authors' equal contribution. The authors acknowledge financial support from by the Spanish Ministry of Science and Innovation (CSO2010-18534)

[ii] The first four waves of the panel have been carried out in collaboration with the CIS (study 2855). Further details on the sample and fieldwork are included in Table A1 of the Appendix. A detailed methodological report (in Spanish) about the survey, including the questionnaires, can be obtained upon request from the authors.

[iii] The number of IMC fails depends on the respondents' fidelity to the study.This, in turn, depends on many other conditionings.The maximum number of IMCs failed during the study was 5, with3 individuals gettingthat score. On average, an individual failed 0.46 IMCs out of the 6.

[iv]We have replicated our estimations using multilevel logistic models (individuals nested in waves) and the results (not shown) are very similar to the ones shown here.