Participatory Approach and Self-reported Outcome: A Deadly Combination for the Impact Evaluation of Development Aid Projects[1)]?

Satoru Mikami

Abstract

The purpose of this study is to examine whether social desirability bias increases when reliance on self-reporting as a measurement strategy is coupled with a participatory approach to intervention in the context of a development aid project. Using the field experiment data carried out in Uganda, where the outcome of intervention was measured by both self-report and objective evidence, this study attempts to separately estimate the portion of bias induced usually by social desirability and the one added by the participatory mode of intervention. The results reveal that the former is consistently substantial and significant, while the latter may occur but is not statistically significant.

Keywords: social desirability, self-report, participatory approach

1. Introduction

Self-reporting, which is the default mode of a questionnaire, can be a breeding ground for social desirability bias that haunts quantitative research across disciplines（Andersen and Mayerl

1）This research was funded by the Japan Society for the Promotion of Science（KAKENHI Grant No. JP18K11815）and the Murata Science Foundation. Ethical approval for this study was obtained from the Institutional Review Board at Ehime University, Makerere School of Social Sciences Research Ethics Committee, and the Uganda National Council of Science and Technology.

2019, Kirkpatrick et al. 2018, King 2022, Vergés 2022, Latkin et al. 2016). This can arise especially when researchers try to measure factors whose possible values include socially stigmatized or applaudable attitudes or behaviors, relying solely on self-reporting by respondents. Of course, as long as researchers do not observe every study subject directly, there is always the possibility that self-reports may not measure what researchers want to measure. Usually, the cost of a possible discrepancy is outweighed by the benefit of collecting a wide range of factors (e.g., generalized interpersonal trust and happiness, to name but a few) from a large sample of respondents with minimum time and cost. However, if the discrepancy is huge, it could ruin the entire research. Various measures, from simple to complicated, are proposed to remedy this bias (Harling et al. 2021, Greenwald and Banaji 1995, Warner 1965, Blair and Imai 2012), but the fundamental solution to this problem is the use of internal and/or external validations, such as biomarkers, third-party testimonies, and circumstantial evidence, albeit they can be costly and, in some situations, unethical (Althubaiti 2016, Gormley et al. 2020, Stark 2019).

Nevertheless, self-reporting may be a sanctuary in which the rights of study subjects to their responses are protected. Researchers, especially when using a quantitative approach, do not directly ask study subjects about the reasons for their attitudes or behaviors. Rather, researchers try to interpret the reasons, or causality, from the association between the responses to different questions. This means that the respondents may not have the prerogative to give meaning to what they have done, even though it is exclusively a private matter without any external economy. Self-reporting, then, may be the last stand in which respondents can defend their dignity against intrusion by researchers. This is the reason that participation-oriented scholars, who usually use a qualitative approach, tend to empathize with the importance of subjective answers from study subjects rather than objective observation when they use, for some reason, a quantitative approach.

One of the data collection and analysis modes that took the argument to the extremes was transformative participatory evaluation. This is one of the approaches to learning from the result of development aid projects and is characterized by 1) the inclusion of program beneficiaries as stakeholders, 2) the extensive participation by the beneficiaries in all phases of evaluation, and 3) the ultimate decision-making on the control of the evaluation process by the beneficiaries (Cousins and Whitemore 1998). The intention is to include the researched, who are usually less powerful in the context of development aid, in the knowledge-creation process. As a corollary of this principle, self-reporting is the indispensable part of this approach in measuring the result of intervention because no one, they believe, can know better than the study subjects about the specific impact of the project on them.

Transformative participatory evaluation is usually, if not exclusively, coupled with the participatory planning of intervention, which shares the same ideological origin and principle. Participatory planning of intervention, also known as participatory rural appraisals, is a project development approach that highly respects the value of study subjects' opinions, regardless of the time it takes in agenda setting, needs identification, operationalization, and implementation of the project. If a development project is planned and implemented through the participatory approach and the result is evaluated through the transformative participatory approach, project beneficiaries play a leading role throughout the process, which is the very reason this combination is widely recommended as a means of empowering aid recipients.

One possible problem entailed in this combination is, however, the aggravated risk of the social desirability bias mentioned above. Since the contents of the intervention are determined by the beneficiaries, it is highly awkward for the same beneficiaries to deny, afterwards, the intended impact of the intervention if there is really no impact at all. In this case, reporting positively about the result is "socially desirable." If the result is measured solely by relying on self-reporting, which is logically legitimate in transformative participatory evaluation, there is good reason to expect a biased result. Even if the combination is ideal for the empowerment of beneficiaries, no check-and-balance mechanism is working that could offer reliable information for the financial sponsors of the project.

To address the above-mentioned concerns, this study tests whether the risk of social desirability bias increases when the intervention is planned participatorily, and its results are measured solely based on self-reports. For that purpose, this study used the actual dataset of a field experiment conducted in Uganda. In this experiment, the household water treatment and safe storage（hereafter HWTS）was facilitated in a participatory or non-participatory manner and whether the beneficiaries were continuing the practice after several months was measured in both subjective and objective ways[2)]. By comparing the discrepancies in the two measurement results between the treatment and control groups, this study attempted to extract additional bias that could be induced by the participatory approach. The results showed that the discrepancy tended to be larger among the treatment groups, but the difference did not reach statistical significance. At the same time, however, the study also indicates that the false report allowed by the self-report could undermine the credibility of the impact evaluation result by showing the contrasting results based solely on the self-report and on objective validation.

The rest of the paper is organized as follows: Section 2 reviews the pre-existing studies that

2）For details about the experiment and the analysis based on the objectively measured result, see Mikami（2022）.

illuminated the situation where social desirability bias occurs and situates the present study. Section 3 details the method employed in this study. Section 4 presents the results, and Section 5 concludes.

2. Literature Review

Responses that are conscious of social desirability, like any other human behavior, are determined by both internal and external factors. Depending on which factors researchers emphasize, studies on social desirability can be divided into two patterns. Researchers who regard social desirability–conscious behavior chiefly as one of the personality traits of respondents typically try first to measure the individual level of sensitivity to social desirability using, for example, the Crowne-Marlowe/Marlowe-Crowne Social Desirability Scale（Crowne and Marlowe 1960, Marlowe and Crowne 1961）. Then, they regress the self-reported responses on the scale together with all other relevant factors. Hebert et al.（2008）, for instance, using 267 participants from four Behavior Change Consortium sites in the United States, pointed out that self-reported answers to fat intake were statistically significantly underreported among women with high scores on social desirability consciousness, regardless of the level of education, while no socially desirable bias was found among male respondents. Mossavar-Rahmani et al.（2013）, focusing on 450 postmenopausal women from the Women's Health Initiative participants in the United States, reported that high social desirability scores tend to result in underreporting energy and protein intake in a statistically significant manner. Adong et al.（2019）, using 751 HIV（human immunodeficiency viruses）positives in Uganda, reported that the social desirability score statistically significantly suppresses the reported level of alcohol consumption among those who admitted recent alcohol use, although the tendency of high social desirability sensitivity to the denial of recent alcohol consumption does not reach statistical significance.

The weakness of the above studies in insisting on the effect of an innate level of sensitivity to social desirability is that individual inclination cannot be manipulated for randomization. As long as the value of the factor cannot be randomized, there always remains the possibility of an omitted variable, which correlates with the proneness to social desirability and, at the same time, has its own influence on the outcome of concern. which is measured through self-reporting.

In contrast, if we consider the level of sensitivity to social desirability to be chiefly determined by external factors, we can use a randomized controlled trial, which is what the second group of studies on social desirability typically conducts. By randomly assigning the condition in which people become more aware of the "norm," we can estimate the impact of social desirability

without bothering omitted variables. Miller et al.（2008）, for instance, studied the impact of social desirability on self-reported levels of fruit and vegetable intake by changing the level of emphasis on the purpose of the survey to 183 female respondents aged 35 to 64, who were randomly sampled from the State of Colorado. They discovered that regardless of the type of question wording（24-hour recall question or general frequency question）, the rate of fruit and vegetable intake inflates statistically significantly in situations where people are more likely to be aware of what is socially desired. Likewise, Kypri et al.（2015）implemented an online randomized controlled trial using a 3,594 Australian student sample to test whether the situation that should induce people's awareness of social norms affects self-reporting on alcohol consumption. The results, however, detected no statistically significant impact of social desirability.

The current study belongs to the latter group of investigations in that it tests the impact of the condition that is expected to compel respondents to match their answers and what is socially desired rather than the truth when there is a discrepancy between the two. HWTS is a socially desirable practice in the rural areas of developing countries, where most households do not have access to running piped water systems on premises. Hence, a measurement based on self-reporting is expected to result in an inflated rate of practice compared to reality. In addition, if aid is provided in a participatory manner in which beneficiaries can choose the method of the HWTS, it should be difficult for beneficiaries to admit openly that they have quit the HWTS because it is cumbersome for them. They are expected to pretend to have been continuing the HWTS to save their faces, which is the norm, even if it is not a reality. In contrast, if the method is imposed on beneficiaries paternalistically by the donor, responsibility lies on the donor side, and beneficiaries should have no additional sense of shame in reporting that they have discontinued the HWTS, even though the pre-intervention level of pressure of the social desirability of the HWTS remains. Below, I detail the procedures of a randomized controlled trial to test the hypothesis.

3. Method

The sample of this study is 192 households in Ntenjeru Sub County in Mukono District, Uganda, where the predominant source of water is a protected or unprotected spring. From among the 53 functioning protected springs, 12 homogeneous springs were purposively chosen, and 16 users were recruited at each water supply point.

The participatory or non-participatory mode of intervention was allocated randomly at each

water supply point. The participatory mode of intervention is characterized by three points. First, beneficiaries are not taught about but facilitated to find out by themselves the cause of and the countermeasure against waterborne diseases using the problem or objective tree method. Second, they can choose their preferred method of the HWTS（boiling or chlorination）. Lastly, during the two-week trial period, they are entrusted to manage the resources for the HWTS（charcoal and clay stove for boiling and chlorine tablet and jerry can for chlorination）without the supervision by the enumerators. In contrast, households in the non-participatory mode are unilaterally taught about the importance of the HWTS and instructed either boiling or chlorination as the method. During the trial period, they were supervised by the enumerators. No leeway was given to this group.

At the end of the two-week trial period and after another 16 weeks, households were investigated to determine whether they were continuing the HWTS based on both self-reporting and the observation of objective evidence, namely the number of E. coli colonies per 1 ml of drinking water. Objective validation was conducted according to the following rules. First, households whose drinking water contained a high level of E. coli were regarded as having discontinued the HWTS, regardless of their self-reports[3)]. This is because an inadvertent recontamination of treated water, which is possible, especially when the boiling method is employed, is usually expected to be limited in scope compared to the usual contamination that naturally occurs in untreated water during storage. In addition, there is a good reason for households to pretend to have been continuing the HWTS when they have discontinued it. For instance, they may erroneously expect to be rewarded for giving good answers to the enumerators. Therefore, households with high contamination levels are regarded as having stopped the practice, even if their self-reports contradict this view. Second, if the observed number of E. coli colonies was small, households that reported that they had discontinued the HWTS were regarded as having stopped. This is because there is no potential incentive for households to pretend to have discontinued the HWTS when they, in fact, continue it. Meanwhile, if the level of contamination is low and a household reports that it is continuing the HWTS, that household is regarded as continuing the HWTS, although there remains the possibility that they are not telling the truth because untreated water can be relatively free from contamination if it is still fresh at the time of observation. In short, this test, whether a household has clean water at the point of use, is a hoop test（Van Evera 1997）, where failing to pass the test eliminates the possibility that the household has actually practiced the HWTS, but passing the test does not by itself affirm that

3）The cut-off point to separate “high” and “low” level of contamination was set to 60 or more E. coli colonies per 1 ml.

the household is continuing the HWTS. Nevertheless, this validation can screen out a substantial number of false self-reports.

Using the two versions of the results measured by self-report and objective validation, the following steps were taken to test the above-mentioned hypothesis. First, I compared the pre-intervention rates of implementation of the HWTS between the treatment and control groups based on both measures. It is expected that no difference exists between the two groups, while significant discrepancies exist between the two results based on different measurement strategies. Fisher's exact test was performed for the difference between the treatment and control groups, while McNemar's test was performed for the discrepancy between the self-reported and objectively validated ratios. Second, I examined the difference between the treatment and control groups in the discrepancies between the self-reported and objectively validated ratios at each observation timing（two weeks and 18 weeks after the intervention）, conditional on the method used for the HWTS（boiling or chlorination）. A larger discrepancy is expected for the treatment group because of the additional psychological pressure induced by the participatory approach. The confidence interval was calculated by bootstrapping the sample with a replacement 10,000 times. Lastly, I showed how an impact analysis of the participatory mode of intervention can be misleading if the endpoint measurement relies solely on self-reporting by comparing it with the results based on the objectively validated endpoint measurement.

4. Results

4.1. Comparison of baseline implementation rates

The prevalence of the HWTS before intervention varied widely, depending on the method used for the measurement. Before the intervention, when the enumerators asked about the practice of water treatment at the point of use, 157 out of 192 households reported that they were basically practicing the HWTS. If their answers were all true, there would be no need to conduct an aid project to promote the HWTS in this area. However, more objective evidence suggests otherwise: Only 15% of households had drinking water completely free from E. coli. The average number of E. coli colonies per 1 ml of drinking water was 69.6.

Figure 1 depicts the rates of practicing the HWTS by treatment arm. Based on the self-report, 77（80.2%）households in the treatment group and 80（83.3%）households in the control group had been implementing the HWTS. However, objective validation reveals that the real rates are much lower. In both groups, only 52（54.2%）households practiced the HWTS. The discrepancies between the two measurement results, 26.0% and 29.2%, were highly statistically signifi-

cant according to McNemar's test ($\chi^2 = 23.04$, $p = 0.000$, $\chi^2 = 26.036$, $p = 0.000$). Regarding the difference between the treatment and control groups, no statistically significant difference was found in terms of both self-reported ($p = 0.709$) and objectively validated rates (zero difference).

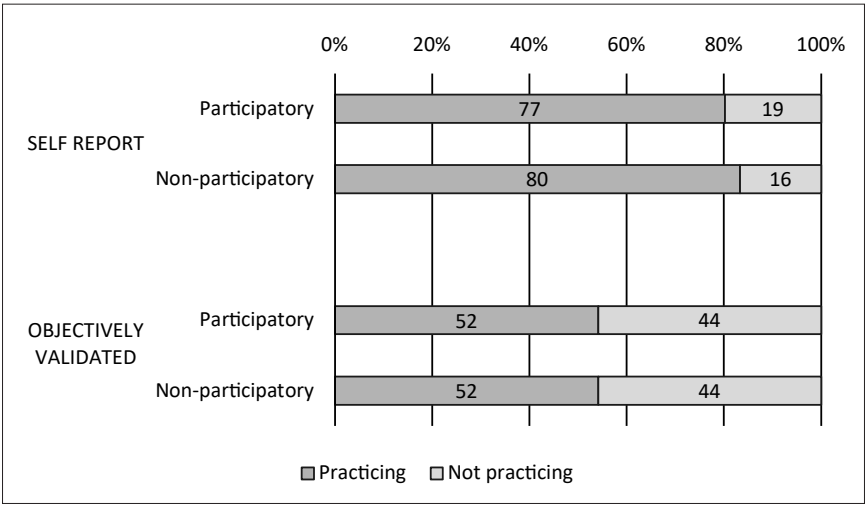

Figure 1. Rates of household who are practicing the HWTS at baseline

4.2. Differences in discrepancies between participatory and non-participatory approaches

Similar to the baseline, self-reports about the continuation of the HWTS after the intervention suffered from the social desirability effect. The two columns on the left in Table 1 list the numbers of households that claimed the continuation of HWST by block (water supply point). Two weeks after the intervention, 189 out of 192 households (98.4%) reported that they always or basically treated drinking water. Another 16 weeks later, 183 households (95.3%) claimed that they had been continuing the HWTS. If we naïvely assume that the self-report reflects reality, we could conclude that the aid project succeeded in boosting the proportion of households that practiced the HWTS from 80% at baseline to almost 100% after two weeks and were able to keep the level at more than 95% for at least another four months. However, the objective indicator suggests otherwise. The objectively confirmed true proportion of households that continued the HWTS after two weeks was 71.4%, and after another four months, it decreased further to 45.8% (the two columns on the right in Table 1). These figures are dauntingly lower but intuitively more plausible than the optimistic estimates based on self-reports.

Table 1. Rates of continued practice of the HWTS by block

	Self-report		Objective validation	
Block	2 weeks later	18 weeks later	2 weeks later	18 weeks later
1	15 (93.75%)	15 (93.75%)	11 (68.75%)	9 (56.25%)
2	15 (93.75%)	15 (93.75%)	10 (62.50%)	7 (43.75%)
3	15 (93.75%)	14 (87.50%)	8 (50.00%)	5 (31.25%)
4	16 (100.00%)	14 (87.50%)	11 (68.75%)	4 (25.00%)
5	16 (100.00%)	15 (93.75%)	10 (62.50%)	7 (43.75%)
6	16 (100.00%)	16 (100.00%)	14 (87.50%)	8 (50.00%)
7	16 (100.00%)	16 (100.00%)	11 (68.75%)	8 (50.00%)
8	16 (100.00%)	16 (100.00%)	13 (81.25%)	7 (43.75%)
9	16 (100.00%)	15 (93.75%)	11 (68.75%)	9 (56.25%)
10	16 (100.00%)	15 (93.75%)	12 (75.00%)	9 (56.25%)
11	16 (100.00%)	16 (100.00%)	14 (87.50%)	9 (56.25%)
12	16 (100.00%)	16 (100.00%)	12 (75.00%)	6 (37.50%)
Total	189 (98.44%)	183 (95.31%)	137 (71.35%)	88 (45.83%)

Table 2 compares the rates of households that continued the HWTS two weeks after the intervention between the treatment and control groups, conditional on the method used for the HWTS. Among the households that chose or were recommended boiling, all households claimed that they were continuing the HWTS, namely 100% for both treatment and control groups. However, the rates dropped to 55.6% and 47.9%, respectively, when the self-reports were confirmed based on objective evidence. The difference between the groups was not statistically significant according to Fisher's exact test ($p = 0.783$). The discrepancy between the rates based on the self-reported and objectively validated results is 44.4% for the treatment group ($\chi^2 = 6.125$, $p = 0.013$), which is unexpectedly smaller than the 52.1% discrepancy for the control group ($\chi^2 = 23.04$, $p = 0.000$). The difference in discrepancies, -7.6%, is not statistically significant at the 5% level according to the bootstrap confidence interval (95% CI: -0.345–0.202, Figure 2, left panel).

In contrast, among the households that chose or were recommended chlorination, the discrepancy is larger for the treatment group, as expected. The rates of continuation based on the self-report were 98.7% for the treatment group and 95.8% for the control group, which decreased to 82.1% and 83.3%, respectively. Again, the difference between the groups was not statistically significant ($p = 1.000$). Meanwhile, the resulting gaps, 16.7% ($\chi^2 = 11.077$, $p = 0.001$) and 12.5% ($\chi^2 = 4.1667$, $p = 0.041$), were statistically significant. The sign of the difference in the discrepancies, 4.2%, is now consistent with the hypothesis, which is not statistically significant (95% CI: -0.088–0.165, Figure 2, right panel).

Table 2. Self-reported and objectively validated rates of continuation of the HWTS (2 weeks later)

	n	Self-reported Survivor	Rate [1]	Objectively validated Survivor	Rate [2]	Discrepancy [1]−[2]	
Boiling							
Participatory	18	18	1.000	10	0.556	0.444	**
Non-participatory	48	48	1.000	23	0.479	0.521	***
				Difference in discrepancy		-0.076	
Chlorination							
Participatory	78	77	0.987	64	0.821	0.167	***
Non-participatory	48	46	0.958	40	0.833	0.125	**
				Difference in discrepancy		0.042	

Note. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

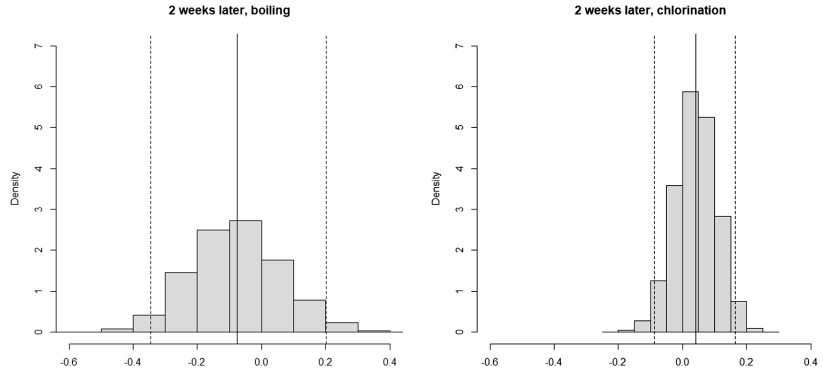

Figure 2. 95% CI of the difference in 2 weeks later discrepancies between participatory and non-participatory approaches based on 10,000 Bootstrapping（left: boiling, right: chlorination）

The results 18 weeks after the intervention are reported similarly in Table 3. All households in the treatment group that chose boiling as the method of the HWTS still claimed that they were continuing the practice, but the more realistic rate based on the objective evidence was only 11.1%. Likewise, almost all households in the control group that were recommended boiling claimed that they were still continuing the HWTS, but the more plausible rate was 18.8% based on the objective validation. The difference in objectively validated rates between the groups was not statistically significant（$p = 0.713$）, while both gaps, 88.9% and 79.2%, between the self-reported and objectively validated rates, were statistically significant（$\chi^2 = 14.062$, $p = 0.000$ and $\chi^2 = 36.026$, $p = 0.000$, respectively）. As expected, the gap for the treatment group was greater than that for the control group but was not statistically significant at the 5% level（95% CI: -0.104–0.274, Figure 3, left panel）.

The same applies for the households that chose or were recommended chlorination. Eighteen weeks later, the self-proclaimed practitioners were 96.2% for the treatment group and 89.6% for the control group, which are inflated rates compared to the objectively validated rates of 60.3% and 62.5%, respectively. The difference between the more realistic rates is not statistically significant（$p = 0.852$）, but the discrepancies between the self-reported rate and objectively confirmed rate are both statistically highly significant（$\chi^2 = 26.036$, $p = 0.000$ and $\chi^2 = 11.077$, $p = 0.000$, respectively）. Lastly, the difference in discrepancies is consistent with the hypothesis but does not reach the conventional level of statistical significance（95% CI: -0.079–0.250, Figure 3, right panel）.

Table 3. Self-reported and objectively validated rates of continuation of the HWTS（18 weeks later）

	n	Self-reported Survivor	Rate [1]	Objectively validated Survivor	Rate [2]	Discrepancy [1]−[2]	
Boiling							
Participatory	18	18	1.000	2	0.111	0.889	***
Non-participatory	48	47	0.979	9	0.188	0.792	***
				Difference in discrepancy		0.097	
Chlorination							
Participatory	78	75	0.962	47	0.603	0.359	***
Non-participatory	48	43	0.896	30	0.625	0.271	***
				Difference in discrepancy		0.088	

Note. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

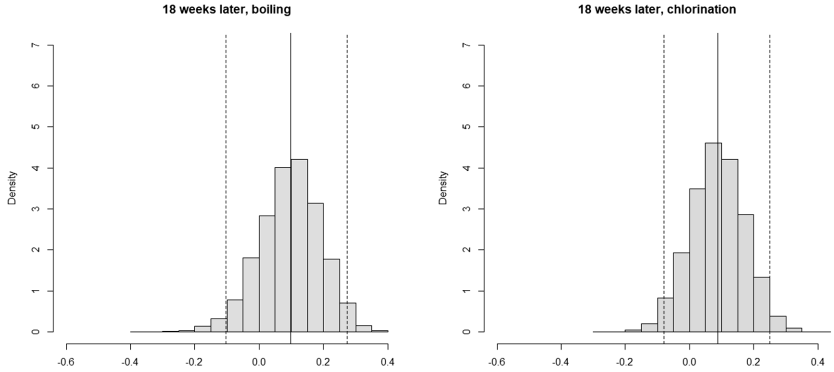

Figure 3. 95% CI of the difference in 18 weeks later discrepancies between participatory and non-participatory approaches based on 10,000 Bootstrapping（left: boiling, right: chlorination）

4.3. Comparison of the results of impact evaluation based on self-reported and objectively validated endpoints

Table 4 reports two different versions of the results of the impact estimation. The top panel refers to the result based on self-reporting, and the bottom panel refers to the result based on objective validation. The impact of the participatory approach on the continued practice of the HWTS with chlorination can be estimated by comparing the survival distributions of the treatment and control groups that used chlorination. When based on the self-report, the point estimates of the

impact of participatory intervention on the 2-week survival rate and on the 18-week survival rate are 2.9 and 6.6 percentage points, respectively. Likewise, the impact of the participatory approach on the continued practice of the HWTS with boiling can be estimated by comparing the survival distributions of the treatment and control groups that used boiling. Again, based on the self-report, the point estimate of the impact of participatory intervention is non-negative both for the 2-week survival rate（0.0）and for the 18-week rate（2.1）. Meanwhile, the estimates of the impact of participation result mostly in the opposite sign when the outcomes are validated with objective evidence. When the method of the HWTS is chlorination, the point estimates of the impact of participatory intervention on the 2-week survival rate and on the 18-week survival rate are -1.2 and -2.2 percentage points, respectively. When boiling was used, the impact on the 2-week survival rate is positive（7.7）, but the impact on the 18-week survival rate is negative（-7.7）. Although all estimates of the impact of the participatory approach do not reach statistical significance, and hence the interpretations of the results converge on “no impact,” the fact that the signs are contrasting suggests the possibility that researchers may reach opposing conclusions depending on which measurement strategy they take if the sample size is large enough to detect any small impact. More importantly, if researchers rely exclusively on the self-report without examining its validity using an objectively verifiable indicator, the statistically significant and counterintuitive impact of boiling（-35.4 and -43.7 percentage points on 2-week and 18-week survival rates, respectively）can be overlooked[4)]. The results did not change after controlling for the block dummies and the covariates that happened to have correlated with assignment status using the Cox proportional hazard model（not reported）.

4）This is the impact of boiling when the non-participatory mode of intervention was taken. The impact of boiling under participatory mode cannot be estimated in this framework because the method（boiling or chlorination）is not randomly assigned but autonomously chosen by the subjects, leading to the intractable problem of omitted variable.

Table 4. Impact estimation of the participatory approach on the continued practice of the HWTS

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	Non-participatory (Chlorination)	Participatory (Chlorination)	Non-participatory (Boiling)	Participatory (Boiling)	Impact of Participation & using boiling [4]-[1]	Impact of Participation when the method is chlorination [2]-[1]	Impact of Participation when the method is boiling [4]-[3]	Impact of using boiling when the mode is not participatory [3]-[1]
N	48	78	48	18				
Self-report								
2 weeks	0.958	0.987	1.000	1.000	0.042	0.029	0.000	0.042
18 weeks	0.896	0.962	0.979	1.000	0.104	0.066	0.021	0.083
Observed	5	3	1	0				
Expected	2.210	3.660	2.270	0.850				
Logrank test					1.970	2.160	0.380	1.970
Objective validation								
2 weeks	0.833	0.821	0.479	0.556	-0.277	-0.012	0.077	-0.354
18 weeks	0.625	0.603	0.188	0.111	-0.514	-0.022	-0.077	-0.437
Observed	18	31	39	16				
Expected	28.060	45.230	21.980	8.730				
Logrank test					14.940***	0.060	0.020	21.040***

Note. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$. Proportions are reported. *P*-values are based on the logrank test.

5. Conclusion

This study aimed to separately estimate the portion of discrepancy between the self-report and the reality caused normally by social desirability and the portion caused additionally by the participatory mode of intervention in the context of development aid. For that purpose, a randomized controlled trial dataset of an HWTS-promoting experiment in Uganda was utilized, in which the two modes of intervention（participatory and non-participatory approaches）were assigned randomly, and the outcome was measured both with self-report and objective evidence. After confirming that there was no significant difference in the size of the social desirability bias between the participatory and non-participatory groups by comparing the pre-intervention rates of the HWTS practice, this study examined whether the post-intervention social desirability bias was larger among those that facilitated the HWTS in a participatory manner. The results showed the expected tendency, but the difference was not statistically significant. Meanwhile, the comparison of the impact evaluation results, one based on the self-report and the other on objective validation, suggested the possibility that the result based on the self-report could be misleading because of the sheer size of bias caused just by social desirability.

Self-reporting is the basic mode of measurement for most questionnaire-based investigations. The measurement strategy works unless a question involves psychological pressure for respondents to respond differently from the truth. However, all too often, the most intriguing part of any study lies in the sensitive questions to which most respondents prefer to answer strategically. Therefore, on the one hand, self-reporting is one of the weak points of the quantitative approach that can ruin the credibility of the results based on the data measured only by self-reporting. On the other hand, from the viewpoint of those who are studied and their defenders, self-reporting can be positively evaluated as a means of protecting the rights of study subjects' privacy. There are kinds of truths, according to this view, that may not be revealed, even for academic purposes. Furthermore, proponents of transformative participatory evaluation go as far as to argue for the need to redefine the truth about the result of intervention, from which the subsequent policy should learn. However, if the intervention is to improve reality, its evaluation needs to remain in the realm of positivism, and the measurement must evolve incessantly to reflect the truth.

References

Adong, J., Fatch, R., Emenyonu, N.I., Cheng, D.M., Muyindike, W.R., Ngabirano, C., Kekibiina, A., Woolf-King, S.E., Samet, J.H., and Hahn, J.A.（2019）. Social desirability bias impacts self-reported alcohol use among persons with HIV in Uganda. *Alcoholism: Clinical and Experimental Research*, 43（12）, 2591−2598

Andersen, H. and Mayerl, J.（2019）. Responding to socially desirable and undesirable topics: Different types of response behaviour? *Methods, data, Analyses*, 13（1）, 7−35

Althubaiti, A.（2016）. Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, 9, 211−2017

Blair, G. and Imai, K.（2012）. Statistical analysis of list experiments. *Political Analysis*, 20, 47−77

Cousins, J.B. and Whitmore, E.（1998）. Framing participatory evaluation. *New Directions for Evaluation*, 80, 87−105

Crowne, D.P. and Marlowe, D.（1960）. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24（4）, 349−354

Gormley, I.C., Bai, Y., and Brennan, L.（2020）. Combining biomarker and self-reported dietary intake data: A review of the state of the art and an exposition of concepts. *Statistical Methods in Medical Research*, 29（2）, 617−635

Greenwald, A.G. and Banaji, M.R.（1995）. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102（1）, 4−27

Harling, G., Bountogo, M., Sie, A., Bärnighausen, T., and Lindstrom, D.（2021）. Nonverbal response cards reduce socially desirable reporting of violence among adolescents in rural Burkina Faso: A randomized controlled trial. *Journal of Adolescent Health*, 68, 914−921

Hebert, J.R, Hurley, T.G., Peterson, K.E., Resnicow,K., Thompson, F.E., Yaroch, A.L., Ehlers, M., Midthune, D., Williams, G.C., Greene, G.W., and Nebeling, L.（2008）. Social desirability trait influences on self-reported dietary measures among diverse participants in a multicenter multiple risk factor trial. *The Journal of Nutrition*, 138（1）, 226−234

King, B.M.（2022）. The influence of social desirability on sexual behavior surveys: A Review. *Archives of Sexual Behavior*, 51, 1495−1501

Kypri, K., Wilson, A., Attia, J., Sheeran, P., Miller, P., and McCambridge, J.（2016）. Social desirability bias in the reporting of alcohol consumption: A randomized trial. *Journal of Studies on Alcohol and Drugs*, 77（3）, 526−531

Kirkpatrick, S.I., Collins, C.E., Keogh, R.H., Krebs-Smith, S.M., Neuhouser, M.L., and Wallace, A.（2018）. Assessing dietary outcomes in intervention studies: Pitfalls, strategies, and research needs. *Nutrients*, 10, 1001. doi: 10.3390/nu10081001

Latkin, C.A., Mai, N.V., Ha, T.V., Sripaipan, T., Zelaya, C., Le Minh, N., Morales, G., and Go, V.F.（2016）. Social desirability response bias and other factors that may influence self-reports of substance use and HIV risk behaviors: A qualitative study of drug users in Vietnam. *AIDS Education and Prevention*, 28（5）, 417−425

Marlow, D., and Crowne, D.P.（1961）. Social desirability and response to perceived situational demands. *Journal of Consulting Psychology*, 25（2）, 109−115

Mikami, S.（2022）. Effect of participation and alignment on the sustainability of development aid output: Evidence from a field experiment in Uganda. *Studies in Comparative International Development*. doi: 10.1007/s12116-022-09374-9

Miller, T.M., Abdel-Maksoud, M.F., Crane, L.A., Marcus, A.C., and Byers, T.E.（2008）. Effects of social approval bias on self-reported fruit and vegetable consumption: A randomized controlled trial. *Nutrition Journal*, 7（18）. doi: 10.1186/1475-2891-7-18

Mossavar-Rahmani, Y., Tinker, L.F., Huang, Y, Neuhouser, M.L., McCann, S.E., Seguin, R.A., Vitolins, M.Z., Curb, J.D., and Prentice, R.L.（2013）. Factors relating to eating style, social desirability, body image and eating meals at home increase the precision of calibration equations correcting self-report measures of diet using recovery biomarkers: findings from the Women's Health Initiative. *Nutrition Journal*, 12（63）. doi: 10.1186/1475-2891-12-63

Prentice, R.L., Sugar, E., Wang, C.Y., Neuhouser, M., and Patterson, R.（2002）. Research strategies and the use of nutri-

ent biomarkers in studies of diet and chronic disease. *Public Health Nutrition*, 5(6A), 977−984

Stark, K.D.（2019）. Applications of innovative lipidomic methods for blood lipid biomarkers. *Journal of Oleo Science*, 68(6) 503−510

Van Evera, S.（1997）. *Guide to Methods for Students of Political Science.* Ithaca, NY: Cornell University Press.

Vergés, A.（2022）. On the desirability of social desirability measures in substance use research. *Journal of Studies on Alcohol and Drugs*, 83(4), 582−587

Warner, S.L.（1965）. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309), 63−69